Conference Proceedings of Science (CPS)
Volume:1 Issue: 1
https://doi.org/10.1007/s44313-025-00108-z
June, 2025

Article OPEN ACCESS

Detection Facial Forensics Based on Deep learning approaches: Comprehensive Literature Review

Israa Adnan Mishkal

Department of Computer Science, College of Science, University of Diyala, Diyala, Iraq.

ABSTRACT

Facial Forensics (FF) is one of the most active research problems in computer vision and digital image forensics, with a wide variety of practical and commercial applications, including identification, access control, and interaction between people with smart devices. However, identifying a face raises serious questions about individual liberties and raises ethical issues. In recent years, important methods, algorithms, approaches, and databases have been proposed for research FF without constraints. There are two approaches namely: the 2D approach has reached a certain level of maturity and has reported very high recognition rates; the 3D approach has been helping to reduce such ambient conditions. It has been proposed as an alternative to the problems mentioned above. The advantage of the 3D dataset is that it is invariant to the pose and lighting conditions, which improves the efficiency of the detection systems. However, the 3D dataset is a bit sensitive to changes in facial expressions. This paper presents FF technologies, currently advanced methods, and future directions. It focuses specifically on the latest database of 2D and 3D facial forensics that utilized for last five years. Furthermore, it focuses on deep learning (DL) approaches due to excellent performances. Also, potential directions for research in facial forensics are presented to provide the reader with a point of reference for topics worthy of consideration in the facial forensic field.

Keywords Facial Forensics (FF), Artificial intelligence (AI), deep learning (DL), face dataset digital forensics.

***Corresponding Author**

Israa Adnan Mishkal
Department of Computer Science, College of Science,
University of Diyala, Baqubah City, Diyala Governorate,3200, Iraq.
Email: israaadnan@uodiyala.edu.iq

© The Author(s) 2025. Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

1. INTRODUCTION

manipulated images and videos are uploaded every day on the Internet; most of them for benign reasons or with malicious intent [1]. Facial manipulation, one of the most common parts, is a very serious problem because faces are not just playing a significant interaction role, but also are widely used in many biological identification and authentication devices [2]. Therefore, reasonable manipulation of facial patterns will seriously undermine trust in digital communication and security applications.

FF encompasses various techniques, including face swapping, the manipulation/animation of facial expressions, and the creation of nonexistent human faces. On the other hand, facial manipulation involves altering facial characteristics such as age, gender, skin color or texture, etc. Utilizing facial analysis techniques, encompassing facial features and additional biometric particulars, such as the eyes, and FF approaches facilitate comparisons with images or video datasets. The contentious nature of this technology arises from allegations of pervasive surveillance, eliciting apprehensions among its adversaries regarding potential encroachments upon data privacy and personal freedoms. Advocates of FF contend that it affords precise, expeditious, and impregnable authentication mechanisms, thereby fortifying defenses against varied forms of fraudulent conduct. So recently, FF, in conjunction with AI methodologies, empowers the identification or verification of individuals based on their facial attributes [2] [3]. FF was appraised at 4.4 billion dollars in 2019 and is anticipated to exceed 10.9 billion dollars by the next four or five years[4].

AI techniques, such as DL algorithms, have emerged as powerful tools for creating manipulating videos and images with incredible realism. Leveraging deep neural networks (DNNs), general adversarial networks (GANs), and other advanced algorithms, facial manipulation deceives viewers by creating near-impossible synthetic media distinguishable from the authentication record. This technology, including its potential has huge impacts on enormous social, ethical, and safety impacts [5].

DL can do significant changes on video and image manipulations such as swap faces or superimpose one person`s face onto another's, creating the illusion that the target individual is saying or doing something they never actually did [6]. This happens by training a DNNS or GANs to learn facial features, expressions, and movements from numerous selecting videos or images, and then applying that knowledge to manipulate the target video or image. The domain of FF has witnessed remarkable progress as a result of advancements in DL methodologies. Firstly, the research paper focuses on FF that has been performed under controlled settings, with both traditional and simplified methods showing exceptional performance. However, contemporary research endeavors have shifted their focus toward addressing the challenges posed by unrestricted conditions. In recent years, DL methodologies have gained significant traction due to their inherent capacity to withstand the diverse variations that can potentially disrupt the recognition processes [7]. This popularity stems from the power demonstrated by distance learning techniques, which enables them to effectively deal with a range of factors that can hinder accurate facial analysis in unconstrained environments. [8].

Furthermore, the difficulty in the FF field often encounters in obtaining robust and dependable datasets for testing and evaluating proposed systems. To accurately evaluate the effectiveness of these systems, it is necessary to access data sets that meet specific criteria. Specifically, the selected datasets should include a large number of individuals and corresponding photographs. In addition, they should adhere to real-world requirements, reflecting the complexities and differences they encounter in practical scenarios. These datasets also should be openly available to the public. The availability of such meticulously constructed datasets plays a critical role in enabling researchers to evaluate and compare the performance of various methods in a comprehensive and unbiased manner[6][8].

To summarize the key findings and developments in the field up to the present date:

1.Provide a comprehensive update overview of FF, highlighting the importance of detecting facial manipulations in various contexts based on DL approaches.

2.Assess and contrast various DL techniques based on their implemented structure and the metrics used to evaluate their performance.

3.Analyze all existing datasets used for FF, and examines the diversity, size, and quality of these datasets, highlighting their implications for training robust DL models.

4.Explore some new directions and future challenges in the field of FF.

Table 1. Techniques for preparing this review

Purpose	▪ To provide a brief overview of existing state-of-the-art techniques and identify potential gaps in FF system detection. ▪ To provide systematic review and structure to the existing state-of-the-art DL approaches that utilized to improve the FF.
Data sources	Elsevier, Springer Link, MDPI and IEEE explore
Query	Methodological approach was designed on using data sources above and the following query strings were used: ▪ Face manipulation ▪ DL for detecting facial manipulation ▪ Deepfakes and Face swap based on DL approaches ▪ 2D dataset for facial forensics ▪ Digital image forensics & DL ▪ 3D dataset for facial forensics ▪ Digital image forensics & DL
Method	We have systematically categorized the literature on Facial forensics detection systems as follows: ▪ Datasets utilized for detecting facial manipulation for last five years. ▪ DL techniques used for detecting facial manipulation for last five years. ▪ CNN Architectures examined for detect facial forensics manipulation. ▪ Open research problems that still persist to this day.
Size	We summarized 145 papers that proposed new FF systems based on DL techniques using the method and query mentioned above from listed data sources.
Study types	The peer-reviewed journal papers, and articles of conference proceedings, were given more importance.

2. Facial Forensics System

2.1. Main stages in FF

Automated FF system entails a series of essential stages, as seen in Figure1. These stages involve:

First stage is the facial preprocessing that refers to the initial stage in facial recognition where raw facial input is processed and prepared for subsequent analysis and feature extraction (FE). This stage involves various techniques and operations to enhance the quality, normalize the appearance, and align the facial images or videos, making them suitable for accurate and reliable FF detection algorithms. During the recognition stage, several operations are typically performed, including [9]:

a)Face Detection: The first task is to detect the presence and location of faces in an image or video frame. Face detection algorithms analyze the input data and identify regions likely to contain faces based on specific patterns, features, or DL approaches.

b) Facial alignment: this technique is applied after detecting the facial region to normalize the facial orientation and pose such as eyes, nose, and mouth. These parts of facial are aligned consistently across different images or frames, minimizing variations caused by head rotation or tilt.

c) Facial Landmark Localization: This includes identifying the main parts of the face, such as the corners of the eyes, nose, and mouth. These landmarks serve as reference points for subsequent operations, such as feature extraction and face normalization.

d) Image Enhancement: This technique used to improve the quality and appearance of facial images. Various image enhancement techniques can be used namely, contrast adjustment, noise reduction, illumination normalization, or other preprocessing operations to enhance visual clarity and remove artifacts.

e) Face Normalization: It is applied to standardize the facial appearance across different images or frames. This technique involves warping operations, scaling, and rotation to align the facial features spatially, making them consistent and facilitating accurate feature extraction.

The second stage involves facial detection and normalization of faces, which aim to approximate the location and alignment of facial regions and detect the presence of faces in an image or video frame, such as the eyes, nose, and mouth. Also, these techniques are applied to standardize the facial appearance by accounting for variations in pose, scale, and illumination.

The third stage used to extract distinctive facial features and then accurately normalize them to enhance the accuracy of the representation. It applied to extract features and normalize the face accurately. Once approximate facial regions are detected, the system extracts meaningful features from these regions. These features capture distinctive information about the facial structure, texture, and appearance. Examples of commonly used features include local descriptors, geometric landmarks, or deep neural network embedding. Furthermore, accurate face normalization techniques are employed to align and normalize the facial features, reducing variations caused by pose, expression, and lighting conditions.

Fourth stage is classification, which encompasses the tasks of authentication (determining if a given face matches a specific identity) and manipulation (assigning an identity to a given face). In the final stage, the system classifies the normalized facial features to perform either authentication or manipulation images or videos [9].

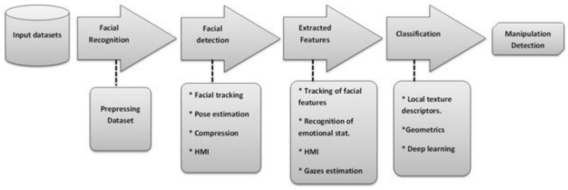

Figure 1. The conventional configuration of an automated FF system

Authentication involves comparing a presented face against a stored template or reference to determine if they match. Manipulation, on the other hand, aims to assign an identity to a given face by comparing it against a database of known identities. This task involves searching for the closest match or ranking potential matches based on similarity scores [9] [10]. In complex environments, various external factors can significantly impact the distribution of intra-face identities (within the same individual) or inter-face identities (across different individuals). These factors can have detrimental effects on the accuracy of FF systems. For instance of these factors include database size, lighting conditions (low or high), presence of noise or blur, disguises, partial occlusion, and other challenging secondary factors that are often common and unavoidable[9].

In the context of this paper, the primary focus lies on FE (and potentially feature selection) and classification. Additionally, the acquisition and detection of faces, which are critical problems in 2D and 3D facial recognitions, are analyzed separately.

2.2 Evaluation Procedures in Facial Detection

As discussed in the previous section, an automated facial detection system can operate in either authentication or manipulation mode, depending on the specific application, as shown in Figure2.

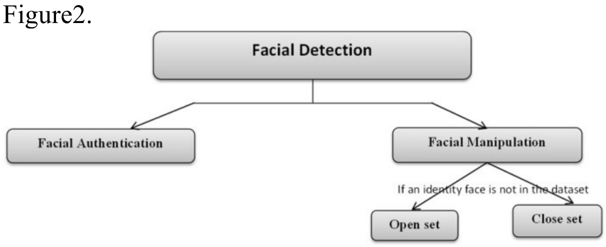

Figure 2. Classification of Diverse Evaluation Protocols in Facial Detection

2.2.1 Authentication mode:

The system assesses an individual's identity by comparing the captured face with the registered models stored in the database. A one-to-one comparison is conducted by the system to determine the authenticity of the claimed identity. Authentication is commonly employed for positive recognition, ensuring that different individuals cannot use the same identity. Receiver Operating Characteristic (ROC) analysis and estimated mean accuracy (ACC) are widely used

assessment protocols for evaluating face authentication systems [9].
Receiver operating characteristic (ROC) analysis evaluates two types of errors: the true accept rate (TAR) and the false accept rate (FAR). TAR represents the proportion of valid comparisons that surpass the similarity score (threshold) accurately [9].

$$TAR = {}^{TP}/_{(TP + FN)} \quad (1)$$

$$FAR = {}^{FP}/_{(FP + TN)} \quad (2)$$

$$ACC = {}^{(TP + TN)}/_{(TP + TN + FP + FN)} \quad (3)$$

Where: TP is true positive for the number of facials that are detected, TN is true negative for the number of facials that are not detected, FP is false positive for the number of non-facials that are falsely detected, FN is false negative for the number of non-facials that are rejected of classifier, ACC is the accurate of detection system.

2.2.2 Manipulation mode:

facial detection system aims to determine the identity of an individual by searching for enrolled model that best matches among all the facial models stored in the database. In the open-set face manipulation scenario, the training set does not include the identities present in the test set. In this case, specific metrics are established to measure the accuracy of the model, such as the false negative rate (FNR) and the false positive rate (FPR). In contrast, the closed-set face manipulation scenario involves using images from the same identities for both training and testing. In this case, a fundamental performance metric used is N-Top, which measures the accuracy of the model by considering whether the correct user identifier is returned within the N-top matches [9].

3. AVAILABLE DATASET FOR FF

In the context of manipulation detection, it is essential to access datasets that encompass a sufficiently large number of images or videos to assess and compare the efficacy of the FF detection system. This accessibility of these datasets to the general public is crucial. This section aims to provide a concise overview of the most suitable and up-to-date datasets for evaluating the performance of FF systems. Additionally, these datasets should be readily available for download or accompanied by a reasonable certification process. This review paper focus will especially be on datasets specifically designed to evaluate 2D and 3D facial approaches. Figure3 presents a summary of these datasets.

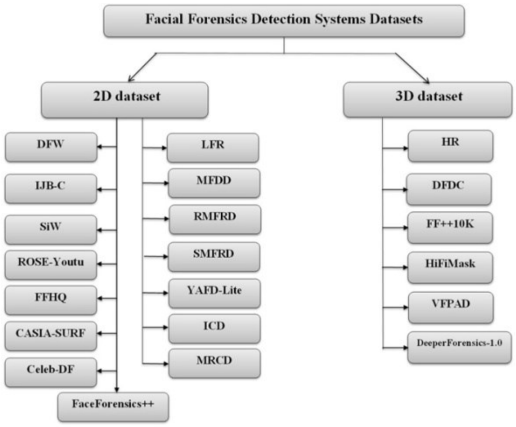

Figure 2. Facial Forensics Database

3.1 2 D-dataset

Data constitute the fundamental cornerstone of any AI model. The acquisition of specific and impartial data from appropriate sources plays a crucial role in constructing a model in both precise and reliable. The following section delves into a comprehensive exploration of the extensively utilized datasets that are openly available for 2D FF to detect face manipulation attacks.

3.1.1 FFHQ dataset: The dataset known as Flickr-Faces-HQ (FFHQ) is comprised of a substantial collection of 70,000 PNG images, each possessing a high level of quality with a resolution of 1024×1024 [15]. This dataset exhibits significant diversity in terms of age, ethnicity, and image background. Furthermore, it offers extensive coverage of various accessories, including eyeglasses, sunglasses, hats, and more (see figure 4). The images were obtained through web crawling of the Flickr platform, which introduces inherent biases associated with the website. To ensure consistency and usability, the images were automatically aligned and cropped using the DLIB library.


Figure 4. Sample of **FFHQ** dataset[15]

3.1.2 LFR Dataset: The LFR (left-front-right) dataset presented by Elharrouss et al. [19] from Qatar University in 2020, addresses the challenge of consistent facial recognition in unconstrained environments. Pose variation represents a major obstacle in achieving accurate face recognition under real-world conditions. To address this challenge, researchers propose a CNN model specifically designed for facial pose estimation. The CNN method is trained using a carefully generated dataset, which combines samples from three widely used reference datasets: Labeled Faces in the Wild (LFW), Celebrities in Frontal

Profile (CFP), and CASIA-Web Face. The dataset includes three distinct facial capture categories: left, front, and right profiles. By leveraging these three categories, a comprehensive dataset is created consisting of images representing 542 unique identities. For each subject, facial images from the left, front, and right views are included in the dataset. It's organized into folders, with dedicated folders for left and right profiles, each containing 10 to 100 facial images. On the other hand, the front folder contains between 50 and 260 images, capturing the frontal view of the subjects. Through the LFR dataset, researchers and practitioners have access to a valuable resource for developing invariant face recognition techniques in real-world scenarios.

3.1.3 Masked Face Detection Dataset (MFDD): This dataset exists amidst the COVID-19 pandemic in 2020 [20]. It is specifically designed to facilitate the training of masked face detection models with high accuracy. It serves as a valuable resource for developing powerful algorithms capable of accurately detecting and localizing masked faces. This dataset contains 24,771 masked face images.

3.1.4 Real-World Masked Face Recognition Dataset (RMFRD) (See Figure 5) includes 5,000 images of 525 individuals wearing masks [20]. In addition, it includes 90,000 images of the same 525 individual people without masks, obtained from various online sources. This dataset aims to simulate real-world scenarios and enable researchers to evaluate and improve the performance of facial recognition algorithms specifically on masked faces.

Figure 5: Sample of **RMFRD** dataset[20]

3.1.5 Simulated Masked Face Recognition Dataset (SMFRD): In addition to RMFRD, the authors utilized alternative methods for fitting masks on widely used large-scale facial datasets such as LFW): In addition to RMFRD, the authors utilized alternative methods for fitting masks on widely used large-scale facial datasets such as LFW[21] and CASIA WebFace [9]. This augmentation technique expands the volume and diversity of the masked facial recognition dataset. The SMFRD dataset encompasses a vast collection of 500,000 facial images, featuring 10,000 individuals, both with and without masks. By incorporating masked and unmasked counterparts, this dataset facilitates practical evaluation and comparison of facial recognition algorithms under masked conditions (as illustrated in Figure 6) [20]. The MFDD, SMFRD, and RMFRD datasets are collected by supporting some institutions namely: National Natural Science Foundation of China, Hubei Province Technological Innovation Major Project, and The Outstanding Youth Science and Technology Innovation Team Project of Colleges and Universities in Hubei Province. They provide researchers with valuable resources to advance the development of robust and accurate facial recognition algorithms specifically for individuals wearing masks, addressing the unique challenges posed by the COVID-19 pandemic [20].

Figure 6. sample of **SMFRD** dataset[20]

3.1.6 YAFD-Lite: This dataset is introduced by L. Liu et al. in 2021[22]. It constructs by extracting children's photos using relevant keywords from a specific social network. This dataset includes 12,181 facial images of children aged (1 to 18 years). To circumvent the need to take multiple photos of the same person at different ages, the dataset was divided into four age groups: 1-4, 5-8, 9-13, and 14-18. In addition, images underwent basic preprocessing such as rotation and resizing to ensure alignment across the dataset.

3.1.7 Longitudinal Indian Child Dataset (ICD): Developed in response to the prevalent problem of child abductions in India. it was introduced by [23] in 2022 and serves the purpose of monitoring and studying the developmental changes of Indian children over time. It consists of 35,484 child face images, comprising 9,475 paired subjects and 7,494 unpaired subjects. Each image in the dataset is labeled with age information ranging from 2 to 19 years, as well as gender information indicating whether the child is male or female. Paired images capture the same subject at different ages, while unpaired images represent situations where only one image of the subject is available. It is worth noting that the dataset consists of 54% of images of male faces and 46% of images of female faces.

3.1.8 The Multiracial Child Dataset (MRCD): MRCD is an extensive compilation of facial images from four distinct racial groups, namely Asian, Black, White, and Indian. This dataset, which is presented in [23] 2022, encompasses a total of (64,965) facial images. The images were acquired through the aggregation of publicly available datasets and web browsing. To facilitate analysis and research, the dataset is organized into five age groups, specifically 0-3, 4-5, 9-12, 13-16, and 17-20. Each racial group within the dataset is well-represented, with Asian comprising (17,211), Black (13,354), White (19,297), and Indian of (15,103) facial images. Table1 present the most recent 2D datasets used for training and testing FF systems. From the table below we see some datasets are suitable for both training and testing purposes, others datasets focus only on training purposes.

Table1: Summary of the most recent 2D datasets used for training and testing FF systems (sorted by years).

2D dataset	Year	Total Images/videos	subjects	Description
[19]	2020	30,000 images	542	The dataset encompasses of LFW, CFP and CASIA-WebFace.
[20]	2020	24,771 images		masked face images
[20]	2020	95,000 images	5000 90,000	525 wearing mask 525 without wearing mask
[20]	2020	500,000 images	10,000	both wearing and without wearing masks
[22]	2021	12,181images		facial images of children within the age range of (1 to 18 years that Divided to age categories: 1-4, 5-8, 9-13, and 14-18
[23]	2022	35,484 child face images		9,475 paired subjects / 7,494 non-paired subjects
[23]	2022	64,965 children images		Asian (17,211), Black (13,354), White (19,297), and Indian of (15,103)

As shown in the table above, the dataset is categorized into two types: images and videos. The images dataset is further divided into adult and younger people. The largest adult dataset is SMFRD, which comprises 500,000 images of 10,000 subjects, with a significant portion of them wearing masks due to the COVID-19 pandemic in 2020 [20]. On the other hand, the second type of dataset is specifically collected for children, with categories based on age, gender, and ethnicity. Additionally, some of datasets exhibit a kind of manipulation such as face swapping, facial reenactment methods, and cutting the eyes, nose, and mouth areas of printed flat or curved faces in encompassing different lighting conditions, poses, and facial expressions (Other details seen in Figure 7 below).

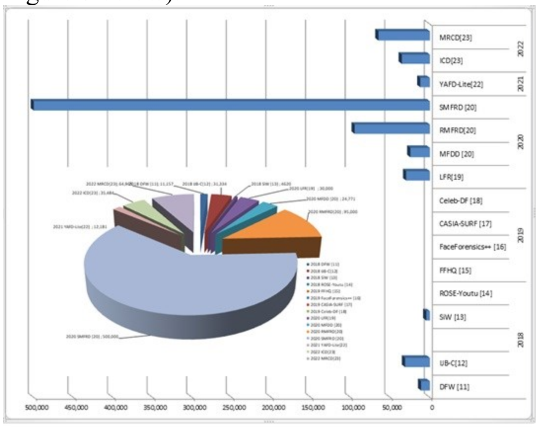

Figure 7: The variety of 2D dataset types over the years

3.2 3D-dataset

The 2D FF datasets have demonstrated valuable in tending to different challenges related to images and video controls [24]. However, FF detection systems need the profundity and spatial data essential to precisely identify and analyze complex controls, such as deepfakes that include practical facial developments and expressions. 3D datasets offer a solution to these limitations by capturing the geometric structure and depth information of the face, providing a more comprehensive representation for forensic analysis [9], [24]. By incorporating different viewing angles, lighting conditions, and facial poses, these datasets enable the development of FF systems that are robust to variations in pose and illumination. As shown in Table 2 and Figure 8, a comprehensive summary and comparative analysis of the facial recognition 3D datasets is presented. The subsequent section delves into an exhaustive examination of the commonly employed datasets that are widely accessible for 3D FF applications.

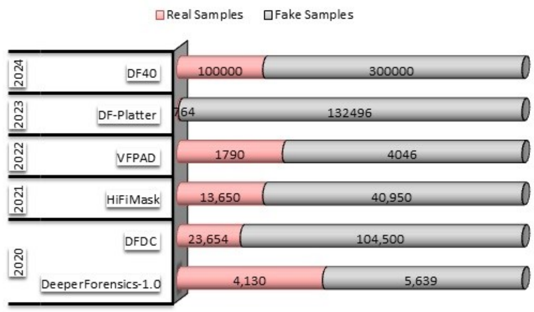

Figure 8: The quantity of 3D datasets classified as Real/Fake videos

Table2: Summary of the most recent 3D datasets used for FF systems (sorted by years)

3D	Year	Samples (Real/ Fake)	subjects	Resolution	Created by
[29]		5,639deepfake 4,130 real	(100): 53 males 47 females	1920 × 1080	Jiang, Liming, et al
[30]	2020	104,500deepfake 23,654 real		A varied range from standard definition to high definition	A collaborative effort involving several organizations and individuals.
[31]	2021	40,950 fake 13,650 real	75		Institute of Automation, Chinese Academy of Sciences (CASIA)
[33]	2022	4046 fake 1790 real	40	(1280 × 720 pixels) with 12-bit resolution	Idiap Research Institute

3.2.1 The DeeperForensics-1.0 dataset

This dataset has made significant contributions to the field of FF by providing a large-scale and diverse video corpus for real-world fake face detection in 2020. It has a benefit in driving research into deepfake detection algorithms and evaluating their performance. The deeperForensics-1.0 dataset includes more than 5,000 original videos and each one is deepfake, resulting in a comprehensive dataset for evaluating the effectiveness of different facial impersonation detection techniques (see Figure 10). This diversity improves the representativeness of the dataset and allows researchers to develop robust detection algorithms for different facial features. Including such variations in the dataset provides a more realistic evaluation platform and promotes the development of algorithms that perform well under real-world conditions. In addition to the original videos, the dataset involves deep fake videos, created using advanced facial processing techniques. These videos mimic real-life fake faces and pose significant challenges to detection algorithms. By combining manipulated videos with original videos, the dataset facilitates the training and evaluation of algorithms on real and manipulated facial data, allowing researchers to evaluate the robustness and generalizability of their methods [29].

Figure 10. the DeeperForensics-1.0 dataset

3.2.2. The Deepfake Detection Challenge (DFDC) dataset

DFDC dataset introduced by WS, Facebook, Microsoft, the AI Communications Integrity Partnership Steering Committee, and academics came together to deliver in 2020. This competition played a significant role in promoting research on deepfake detection techniques and in the process developing algorithms. Vigorously combat the spread of manipulated media content. This dataset provides researchers with a comprehensive and diverse collection of real and manipulated videos, enabling evaluation and benchmarking of deepfake detection algorithms. In contrast, deepfake videos are created using advanced facial processing techniques, simulating manipulated faces like real ones. In inclusion, this data allows researchers to develop and evaluate algorithms that includes all authenticities [30].

3.2.3 The HiFiMask

It is the largest dataset currently available for Face Mask Presentation (PAD) attack detection. It was proposed by Arashlu [31] in 2021. It includes 54,600 videos from 75 subjects with three different skin tones. Each individual wears three high-definition masks made of transparent, plaster, and resin materials. The dataset covers six complex scenes. These are white light, green light, periodic three-color Light, and outdoor settings. Preprocessing steps are applied to remove irrelevant background regions and sample 10 frames at regular intervals after face detection. The dataset serves as a valuable resource for researchers and practitioners, facilitating the development and evaluation of algorithms and models for detecting manipulation attacks [31].

3.2.4 In-Vehicle Face Presentation Attack Detection (VFPAD) dataset

It includes 4046 real-life recordings from 40 subjects and 1790 attack demonstration videos using 89 incidental attack artifacts (PAIs). It was introduced by the IDEAB Research Institute in

2022 [31]. The dataset, captured with a near-infrared (NIR) camera and vehicular NIR illumination, covers a range of presentation attacks, including photo printouts, digital displays, and 3D masks. With its diverse topics and attack scenarios, the VFPAD dataset serves as a valuable resource for the development and evaluation of face-manipulation attack detection systems tailored to vehicle settings. It enables researchers to comprehensively analyze the effectiveness and robustness of detection algorithms in real-world conditions. Table 2 is a summary of the datasets commonly used in the literature for FF problems from 2018 to 2022. Datasets are categorized into manipulated and real videos. The significant number of manipulated videos is primarily contributed by DFDC with varied range of resolutions [31].

4. DEEP LEARNING APPROACHES IN FF system:

DL has emerged as a powerful technique in the field of FF systems, enabling advanced analysis and detection of manipulated or authenticated facial images and videos [9]. With the increasing prevalence of image and video editing tools, it has become crucial to develop robust and accurate methods to identify and authenticate facial content. Therefore, this paper conducted an extensive literature review to identify the most effective and up-to-date DL techniques specifically employed in FF systems. Table 3 summarizes the most popular DL approaches that used to detect FF recent six years. These techniques have been tailored to address the specific requirements of facial problems, although they may have limitations in certain scenarios. DL approaches can be separated into the models employed for image detection and the models utilized for video detection.

4.1 Image Detection: Several approaches have been extensively investigated to detect the FF system for manipulated images using deep neural networks (DNNs). The authors [35] propose a neural network-based approach for detecting fake images. This methodology focuses on analyzing statistical characteristics of images to enhance the identification of artificially produced facial images. The paper [36] introduces another strategy based on deep Convolutional Neural Networks (DCNNs) to identify fraudulent images. This approach utilizes a DL network to extract facial attributes from face identification networks. In [37], the authors utilize image pre-processing techniques such as Gaussian blur and Gaussian noise. These techniques aim to improve the mathematical similarity between authentic images and manipulated one imitation at the pixel level. By a playing previous, the scientific classifier can capture more intrinsic features and exhibit improved generalization capabilities compared to earlier FF techniques [38].

Moreover,[39] introduces a novel methodology for deepfake detection by leveraging the self-consistency of local source features. These features represent spatially local and content independent details of images. A convolutional neural network (CNN) approach is employed to extract these features using a novel representation learning methodology, whereby the features are encoded as down-sampled feature maps referred to as pairwise self-consistency learning. The objective of this method is to impose penalties on feature vector pairs that pertain to regions within the same image and demonstrate insufficient cosine similarity scores. However, it is worth noting that this approach may possess limitations when confronted with counterfeit images generated by technologies that directly produce the entire image and maintain consistent features across all points within the image. More studies demonstrate in Table 3.

Therefore, DL typically involves three key stages [40] namely: facial pre-processing, deep features extraction, and deep features extraction for facial matching. The more details on these stages are:

4.1.1 Facial pre-processing:

It plays a crucial role in enhancing the performance of DL-based FF systems, particularly in uncontrolled environments with varying illumination, pose, and facial expressions. Many recent studies have demonstrated that such variations can have a negative impact on DL performance, underscoring the necessity of facial pre-processing techniques to mitigate these effects [41]. To address these challenges, a DL approach is employed for pre-processing, wherein deep convolutional neural networks (CNNs) are trained to be pose-invariant by generating additional images with different poses from a single image. This approach is adopted to overcome the limitations of collecting a large number of images for the training stage. Some studies utilize data augmentation techniques such as photometric transformations [42] and geometric transformations[43][44].

Others involve the reconstruction of a 3D face model and subsequent generation of 2D images with various poses [45] [46]. Additionally, other studies employ CNN models to directly generate 2D images without the need for 3D modeling [47] [48]. Furthermore, it has been demonstrated that combining features extracted from an input image and another image, even if the latter is not part of the training dataset, can successfully synthesize realistic face images [49]. Generative Adversarial Networks (GANs) have also been employed for this purpose [50] [51] [52] [53].

Another pre-processing approach aims to obtain a canonical view of the face image by utilizing images captured from various angles in uncontrolled environments. Stacked Autoencoders (SAEs) [54], CNNs [55], and

GANs[56] have been utilized to generate frontal face images by combining patches from multiple images with different angles. Recent studies have proposed methods such as face racialization based on CNNs [57] and adaptive pose alignment techniques that dynamically learn alignment templates using facial poses[58]. Furthermore, researchers have introduced an illumination robust pre-processing method [59] that effectively removes soft and hard shadows while preserving identity-related information. Deep FE constitutes a vital stage in the FF system. The selection of appropriate features significantly impacts the overall performance of the chosen DL approach.

4.1.2 Deep features extraction:

This stage is a critical aspect of designing a deep learning approach for FF. The selection of network architecture and loss function plays a pivotal role in this process. The CNNs' architecture can be categorized into essential networks and multiple networks [60]. For example, AlexNet[61], VGGNet [62], GoogleNet [63], ResNet [64], and SENet [65] are types of CNN architectures that have gained significant attention from researchers due to their outstanding performance in solving computer vision problems (more details find in the next section). These architectures, along with their variations, have also been applied to FF systems. Additionally, networks with multiple structures have been proposed for multi-task learning, including facial recognition [66, 67].

The loss function is playing a crucial role in training deep CNNs in FF. It has been observed that some active networks such as the softmax alone are insufficient for effectively separating features, as within-class variations tend to be larger than between-class variations. To address this issue and enhance feature discriminability, alternative loss functions have been proposed, such as Euclidean-distance-based loss[68] [69], triplet loss [70], angular/cosine-margin-based loss[71], and various adaptations of the softmax loss [72] [73].

Deep FE in one-shot FF system becomes recently particularly challenging, where only a single or a few images of a subject are available in the dataset [74, 75, 76, 77,78]. This remains an open research problem, as representing data variance with limited samples is difficult. Several approaches have been proposed to tackle the one-shot FF problem. For instance, [74] intermediate deep attribute representations obtained by fine-tuning a DCNN for specific attributes (e.g., gender and face shape) have been utilized, yielding better performance compared to purely face-based feature representations. Another method employs [75] a regularization function in conjunction with the cross-entropy loss function, introducing a new underrepresented-classes promotional loss. This approach achieved outstanding results in the MS-Celeb-1M Low-shot Face Recognition Challenge at ICCV 2017 [77]. In a different approach, a generative adversarial network (GAN) is employed to synthesize meaningful data for one-shot classes, enabling the classifier to better learn these classes [76]. This synthesis is achieved by adapting the data variances from other classes with a larger number of samples.

4.1.3 Deep Features Extraction for Facial Matching

FF can be classified into two main categories namely: face authentication and face manipulation. In both cases, a predefined group of known subjects is initially enrolled in the system, known as the gallery, while during the testing phase, a novel subject, known as the probe, is presented. Following the training of deep networks on extensive datasets under the guidance of an appropriate loss function, each test image is subjected to the networks to acquire a deep feature representation. By employing cosine distance or L2 distance, face authentication determines the similarity between the gallery and the probe, thereby establishing whether the two images correspond to the same subject. Conversely, face manipulation calculates to discern the specific identity of a probe face. Moreover, supplementary techniques are introduced to post process the deep features, ensuring efficient and accurate face matching. These techniques encompass metric learning, sparse-representation-based classifier (SRC), among others [9] [40].

4.2 Video Detection: The identification of fraudulent videos poses challenges for video detection methods due to the degradation of frame content caused by video compression [79]. The compression process leads to the loss of frame data, rendering traditional DL algorithms designed for image identification inadequate for video analysis. Moreover, the temporal characteristics of videos exhibit variations across frames, making it difficult to detect fake content. Researchers have tackled this problem by leveraging the spatiotemporal attributes of video streams to identify deep fakes and uncover subtle face manipulation artifacts introduced during frame-by-frame video editing [80]. However, a limitation arises from the reliance of DL algorithms on online face photos, resulting in a scarcity of images featuring closed eyes. As a consequence, deep fake algorithms struggle to generate realistic blinking patterns characteristic of genuine videos.

To address this limitation, [81] researchers have proposed a method that involves extracting eye regions from videos and employing long-term recurrent convolutional networks to predict dynamic states, enabling the distinction between authentic and fake videos[82].

Another approach [83] utilizes DL techniques to detect deep fakes, with evaluation conducted on datasets such as UADFV [84] and Deep fakeTIMIT. Notably, this approach eliminates the need for pre-generated deep fake videos as negative examples during training. Instead, negative instances are generated dynamically by manipulating the face region of an original image, leading to significant savings in time and computational resources compared to previous methods.

Although embedding techniques have shown some effectiveness in addressing pose variations and occlusions in still image-based facial recognition (SIFR)[85] ,[86], their applications in video detection are limited. However, certain methods initially developed for SIFR, such as DeepID2 [87], C-FAN [88], VGGFace [89], DeepFace [90], and FaceNet [70] have demonstrated promising performance in video detection. In the case of C-FAN [88], the authors utilized a CNN-based model trained on SIFR and integrated a quality value estimation module [88], which effectively consolidated deep feature vectors into a single vector for FF in videos.

Another approach is the Neural Aggregation Network (NAN) that was introduced by Yang et al. [91]. NAN takes a collection of facial images or facial videos as input and generates a compact and fixed-dimensional feature representation. It employs a DCNN for features extraction and utilizes a Siamese neural aggregation network with minimized average function loss for facial detection. For manipulation, a fully connected layer followed by softmax and loss function is employed. The study[92] proposed a novel method called Face and Body Association (FBA) for video detection. FBA involves using a retrained YOLO detector for face detection and a single DNN with ResNet-50 architecture for detection.

Recent advancements in video detection include the application of DL techniques, such as ADRL and automatic face aging [93]. The studies [94] [95] proposed real-time video models based on deep convolutional neural networks (DCNN). The study [96] introduced a dependency-aware attention control (DAC) model, which uses DL to make sequential attention decisions in image embedding.

5. DEEP CNN FOR FACE DETECTION BASED ARCHITECTURES:

In this section, we present several deep FF methodologies based on Convolutional Neural Networks (CNNs) architectures, commonly trained in a supervised manner. Various studies about endeavors have attempted to prepare profound CNNs for FF purposes. These approaches encompass training multi-class classifiers that effectively discriminate between diverse facial identities during the training phase, often employing techniques such as the Softmax activated layer. Additionally, some methods prioritize the acquisition of discriminative deep facial features.

Some techniques concentrate on extracting features from different facial regions, utilizing multi-CNN architectures, while others concentrate on capturing appearance variations from non-frontal facial images. Many studies embrace concepts from metric learning, combining various activation layers or employing effective methodologies for loss function optimization. Furthermore, some investigations explore the utilization of appropriate activation functions. Table 4 summarizes many studies that are organized according to their respective architectural designs.

Several DL approaches for image/video detection are described in Tables 3 and 4, summarizing the recent DL methodologies used for FF detection. As shown in these tables, the most commonly utilized DL technique is CNNs with various architectures. In 2018, the ShallowNet was used to detect fake faces generated by GANs, achieving the highest accuracy with higher resolution image datasets, but it's the performance decreases with smaller image sizes. Other CNNs architectures, namely VGG16, VGG19, ResNet, DenseNet, NASNet, and XceptionNet, have been presented for detecting facial forensics and achieved 80% accuracy in extracting facial attributes. However, they are limited by small images and low resolution. Another approach used LRCN to understand the temporal aspect of eye-blinking, achieving 90% accuracy using CEW dataset. Meso-4 and MesoInception-4 CNN models were proposed to detect FF using two types of datasets, namely Deepfake and Face2Face, achieving 96.9% and 98.4% accuracy, respectively, based on temporal correlation in eye blinking. Another presented approach based on CNN called Area Under Curve (AUC) achieved 93.8% accuracy for images and 97.8% accuracy for videos on UADFV dataset, but it performed less effectively in the presence of changes in illumination, head motions, and face occlusions. Conv-LSTM is another proposed technique based on CNN. It utilized 80 frames and Convolutional LSTM, achieving 97.15 on the HOHA dataset.

In 2019, several researches have been developed to detect and analyze FF based on DL. A Gaussian blur and Gaussian noise DCGAN-CNN achieved 95.45% accuracy in improving the mathematical similarity between real and manipulated images at the pixel level. ResNet, DenseNet, and RCNN models achieved accuracies 94.35%, and 96.3% for detecting manipulated face images/videos. The C-FAN model achieved high accuracy using the IJB-A and YTF datasets. The Y-shaped Autoencoder CNN achieved 93.01% accuracy on FF++ dataset after resizing images to specific size. In 2020, ResNet-50 pre-trained with ImageNet was proposed, utilizing common artifacts of CNN-generated images to achieve 92.9% using a large number of fictitious images produced by a high-performing unconditional GAN model. ADDNet-2D and ADDNet-3D approaches were used to detect FF, but they achieved the lowest accuracies of 76.25% and 65.50%, respectively, on WildDeepfake dataset due to the temporal information contained in deepfake face sequences. The ResNet50 model using the attribution-based confidence metric achieved 96% accuracy in reducing the dimensionality of the input space for feature detection. Other techniques based on CNN were also presented in 2020. One of them achieved 87.62% and 91.07% on images and videos, respectively, with higher accuracies achieved for shorter videos and mid-size ROIs. The other technique was tested on different types of datasets but showed reduced performance when an image was entirely synthetic or encounters low resolution. VGG and ResNet achieved high accuracy using small images selected of CelebA but were not tested on other datasets. In 2022, the ConvLSTM CNN model achieved 99% accuracy in detecting fraud in videos. However, the model failed to capture the temporal behavior of the forged content over time.

6.DISCUSSION AND RESEARCH AREA CHALLENGES:

The FF detection system requires various methods to access and mitigate the harmful impact of image/video manipulations, commonly known as deepfakes. In this study, we present the most effective existing deepfake detection techniques based on DL. These techniques have proven to be effective in identifying fake images/videos, especially when applied to big datasets that include high-quality deepfakes. However, some of these techniques have struggled to handle out-of-domain data and real-world cases, which are primary concerns of detection algorithms. Furthermore, certain limitations arise from the training process and the use of public datasets that are used for validation (mentioned below). The development of DL (GANs) has facilitated the creation of relatively deepfakes. Moreover, detection research in this field inadvertently fuels the improvement of approaches capable of evading detection. Enhanced detection and correction techniques in deepfakes led to the emergence of sophisticated, hard-to-detect manipulated facial forensics. The existence of such manipulation cases is not only difficult to detect deepfakes but also to demonstrate their authenticity. This is a task that most research studies struggle with due to insufficient transparency in their decision-making process. DL approaches often suffer from some challenges namely as:

1) Modeling aging patterns and securing data privacy [136]: This is an ongoing challenge due to various factors such as a limited training dataset, different lighting conditions, expressions, environmental factors, and genetics. Most studies often focus on adult aging and lack sufficient child-aged face images. Additionally, securing data privacy issues requires appropriate regulations as this technology advances. A challenge is to develop additional modules within DL algorithms to ensure data privacy and confidentiality

2) Face recognition in video surveillance has gained significant popularity and is applied in diverse domains, particularly in the field of video surveillance [139]. In this context, the performance of facial recognition systems is heavily dependent on the conditions of image acquisition, particularly when there are changes in posture, and the acquisition techniques themselves may introduce artifacts. Specifically, camera focus issues can result in image blurring, low resolution, compression-related errors, and block effects. The primary challenge faced by face recognition systems in video surveillance scenarios is the accurate identification of individuals from photographs captured by surveillance cameras, which often exhibit characteristics such as blurriness, low resolution, block artifacts, and varying facial poses. Addressing this challenge remains an unresolved problem, necessitating further research efforts.

3) Face anti-spoofing and building a training dataset: Despite the superiority of deep neural networks over humans in face recognition tasks, it has been observed that DL networks are more susceptible to deception. To address this issue, efforts have been made to develop more robust deep networks, as demonstrated by the organization of the recent deep-fake challenge [140]. Enhancing the resilience of face recognition systems against spoofing attacks is an intriguing research direction.

In recent years, numerous methods based on deep neural networks have been proposed for face anti-spoofing [141] [142] [143], showcasing successful performance against various types of spoofing attacks. However, with the emergence of new forms of spoofing attacks, there is a need for zero-shot face anti-spoofing approaches in the future. Additionally, research is actively focused on improving and validating the robustness of deep neural networks against adversarial and semantic attacks. Moreover, finding a dataset requires appropriate regulations as this technology advances [140]. One challenge is to develop additional modules within DL algorithms to ensure successful performance against various new kinds of spoofing attacks.

4) Reducing computational time:

Recently, the primary objective for researchers is to develop more efficient cost functions to reduce computational time [144]. They have explored the possibility of combining existing loss functions such as Softmax loss and Centre loss [145]. Additionally, they have experimented with using different cost functions in different layers time [144].

7. CONCLUSION

This study comprehensively explores the new developments in FF detection based on DL approaches. The findings of this study unequivocally demonstrate a significant increase in research efforts within the field of FF detection, particularly over the past six years. This notable progress can be attributed to the utilization of DL approaches, which have shown superior performance when compared to widely used computer vision methods. Furthermore, a multitude of datasets are available for research and commercial applications. The comprehensive attributes and evaluation protocols of these datasets are also presented in this review study. The initial section of this survey presents an overview of FF detection systems, tracing their historical development. Subsequently, it offers facial detection techniques, in addition to providing a summary of widely used facial datasets (both 2D and 3D) from the period 2018 to 2022, which are employed for training and testing purposes in the field of FF. Detailed examinations of image-based and video-based FF methods are conducted, accompanied by comparative tables that facilitate methodological comparisons. Moreover, this study showcases the deep CNN architectures utilized in the FF detection, accompanied also by comparative tables organized according to their respective architectural designs. Moreover, Several DL approaches for image/video detection are presented in this study. As the result the most commonly utilized DL technique is CNNs with various architectures.

Table 3. The most popular DL approaches for FF

Ref.	Yea	Methodology	dataset	Accuracy	Key Features & weakness
[35]	2018	a shallow (CNN) architecture	CelebA PGGAN	99.99%	Fake human faces created by GANs-Generated Image achieved the best with images at higher resolutions. The performance reduces when using small images sizes.
		Different CNN-based models	Photoshopped dataset	74.9%	Fake human faces created by Human-Created Image perform the best with 128×128 images sizes.
[36]	2018	DCNNs	CelebA	80%	DL network to extract facial attributes from face identification networks. The limitation points are small size and low resolution.
[81]	2018	Long term recurrent-CNN(LRCN)	CEW	90%	LRCN to understand the temporal patterns of eye blinking based temporal correlation
[79]	2018	CNN	Deepfake Face2Face	96.9 98.4	The key feature is based on blurriness, where the eyes exhibit the highest level of intricacy in authentic images, while the background assumes this role in manipulated images due to the facial dimension reduction.
[83]	2018	CNNs	UADFV	Image: 93.8% Video: 97.8%	The proposed system utilizes pre-generated deep fake videos as negative examples during training. However, the system's performance is reduced when there are changes in illumination, head motions and face occlusions.
[92]	2018	DNN with ResNet-50	JANUS CS3	80.4%	YOLO detector for face detection. The weakness of this system lies in the fact that the face and/or upper body remain similar across different scenes and shots, relying heavily on a single feature.
[97]	2018	Convolutional LSTM	HOHA	97.1%	automatically detect deep fake videos
[37]	2019	DCGAN-CNN	CelebA-HQ	95.45% based prepressing operation	improve the mathematical similarity between authentic images and manipulated one imitation at the pixel level

[80]	2019	RCNN	Deepfakes, Face2Face FaceSwap	96.9% 94.35% 96.3%	Choosing features of the faces are(corners of the eyes, the tip of the nose, and corners of the mouth)
[88]	2019	CNN	IJB-A YTF	93.99% 96.50%	The C_FAN effectively consolidated deep feature vectors into a single vector. However, due to limitation of using YTF and IJB-A datasets instead of real-world video, the C-FAN performance is relatively small.
[98]		CNN	FF++	93.01%	Use the multi-task learning strategy and videos. It detects all images after resizing to specific size.
[27]	2020	ResNet-50 pre-trained with ImageNet	mAP	92.9%	Using a large number of fictitious pictures produced by a high-performing unconditional GAN model, i.e., GAN, based on some degree of JPEG compression, blurring, and resizing images.
[99]		ADDNet-2D ADDNet-3D	Wild Deepfake	76.25% 65.50%	ADDNet-2D achieves lowest accuracy in detecting wild deepfakes. Additionally, indicates that ADDNet-3D is not sufficiently sensitive to small deepfake modifications, likely due to the temporal information contained in deepfake face sequences
[100]		ResNet50 model pretrained on VGGFace2	COHFACE YOUTUBE VidTIMIT	96%	The feature detection using the ABC measure to reduce the dimensionality of the input space.
[101]		CNN	Deep Fakes faces Deep Fakes videos	87.62% 91.07%	The biological signals are extracted from portrait videos and employed as an implicit indicator of authenticity. The higher accuracy only achieves in the wild for shorter videos, and for mid-size ROIs.
[102]		VGG & ResNet	dataset consists of 10,000 images: 5,000 real randomly sampled from the CelebA	99.7% & 93.2%,	Introduce adversarial perturbations to enhance deep fakes and fool deep fake detectors. This methodology does not test on other data.
[103]		CNN	DF F2F FS NT	99.17% 98.57% 98.21% 98.13%	Find the border between the target and original faces. Its performance reduces when an image is entirely synthetic or encounters low resolution.

[104]		S-MIL CNN	FF++ Cele b-DF DFD C FFP MS	99.84 % 99.23 % 83.78 % 90.71 %	The experiment is evaluated based on the average accuracy of fake and real testing videos. This evaluation method takes into account both the spatial level and along the temporal dimension. However, since manipulated faces are made separately, the sequences along the temporal dimension may not ehibit the same smoothness as real face sequences.
[105]		CNN	FF++	99.65	Utilize the Deepfakestack deep ensemble learning approach, which relies on the use of a large model size. However, it is crucial to construct the model carefully to avoid overfitting.
[106]		A new CNN model, namely SC net	GFF	93.77 %	It uses wrinkles to detect manipulation. The performance of SCent diminishes when the low-level features extracted by the first layer are sparse, as this adversely affects the subsequent layers' FE process.
[39]	2021	CNN	FF++ CD2 DFD C-P	99.79 % 99.98 % 94.38 %	The features are encoded as down-sampled feature maps referred to as pairwise self-consistency learning.
[107]		CNN	GGF I FFM I	99.94 % 99.98 %	the Multilevel Facial Conceptual Extraction and Cascade Attention Mechanism. The performance of it depends on both LAMs and SAM together.
[144]		CNN	FRA UD1	99%	some weaknesses that appear in this system include the lack of intra-video frame detection, the type of the performed forgery, and forgery localization in video frame
[145]	2022	CNN	DFD C FF++ Cele bDF	99.26 % 98.61 % 99.13 %	The system's performance has some limitations when dealing with data that contains noise and blurring attacks. Where it achieved 96.89% for DFDC, and 96.62% for the FaceSwap, and Face-Reenactment deepfakes. Additionally, the model fails to capture the temporal behavior of the forged content over time.

Table 4. Summary and comparison of Investigations Deep CNN-Based Architectures

Ref.	r e a r h	Description	dataset	Accuracy
[90]	2014	9-layer deep neural Network training over 4000 identities	LFW	97.35%
[109]	20 14	the Deep hidden IDentity features (DeepID) training 10,000	LFW	97.45%
[87]	20 14	Deep IDentification verification features (DeepID2)	LFW	99.15%
[110]	2015	DeepID3	LFW LFW rank-1	99.53% 96.0%
[86]	2015	deep convolutional neural network (DeepID2+ nets)	LFW YouTube	99.47% 93.2%
[111]	2015	Semantic bootstrapping method training 100,000 images	LFW Rank-1 DIR False Alarm Rate(FAR)	72.3% 46.3% 1%
[85]	20 15	multi-patch deep CNN with deep metric learning	LFW	99.77%
[62]	2014	deep convolutional networks (up to 19 weight layers) includes images of 1000 classes	ILSVRC-2014	
[89]	2015	a triplet loss function and face alignment	VGGface dataset	
[46]	2016	specific appearance variations, including shape, pose, and expression using a 3D generic face.	Softmax CASIA WebFace (494 k, 10 k)	98.06
[112]	2016	Range Loss	LFW YTF	99.52% 93.70%
[71]	20 16	L-Softmax	LFW	98.71%

[113]	2017		Noisy Softmax	MNIST CIFAR10 CIFAR10+ CIFAR100 LFW FGLFW YTF	0.33% 7.39% 6.36% 28.48% 99.18% 94.50% 94.88%
[63]	2015			DistBelief	43.9%.
[70]	2015		128-dimensional representations from deep convolutional networks, trained on 200-million facial images by utilizing a triplet loss function	LFW YouTube Faces(YTF)	99.63% 95.12%
[114]	2020		an adaptive fusion of softmax loss and center loss	LFW YouTube Faces (YTF)	
[115]	2016		center loss	(LFW) YouTube Faces (YTF)	99.28% 94.9%
[116]	2017	LeNet	center invariant loss function	LFW YTF	99.12% 93.88%
[117]	2019		feature transfer learning (FTL) that adapts under-represented (UR) classes' feature distribution	MS-Celeb-1M (4.8 M, 76.5 K)	99.55
[118]	2017		L2-Softmax	ResNet-101 1 L2-Softmax MS-Celeb 1M (3.7 M, 58 k)	99.78
[119]	2017		Marginal Loss	ResNet-27 1 Marginal Loss MS-Celeb 1M (4 M, 82 k)	99.48
[120]	2017		NormFace	CASIA WebFace (494 k, 10 k)	99.19 ± 0.008

[121]	2017	COCO Loss	COCO Loss MS-Celeb 1M (3 M, 80 k)	99.86
[122]	2017	Von Mises-Fisher	vMF Loss MS-Celeb-1M (4.61 M, 61.24 K)	99.63
[72]	2018	SphereFace	A-Softmax CASIA WebFace (494 k, 10 k)	99.42
[123]	2018	Ring Loss	Ring Loss MS-Celeb-1M (3.5 M, 31 K)	99.50
[75]	2018	MLR	CCS Loss MS-Celeb-1M (10 M, 100 K)	99.71
[124]	2018	Cosface	Large Margin Cosine Loss CASIA WebFace (494 k, 10 k)	99.73
[125]	2018	AM-Softmax	AM-Softmax Loss CASIA WebFace (494 k, 10 k)	99.12
[126]	2018	Light-CNN	Softmax MS-Celeb-1M (5 M, 79 K)	99.33
[127]	2019	A_nity Loss	A_nity Loss VGGFace 2 (3.31 M, 8 K)	99.65
[128]	2019	ArcFace	ArcFace MS-Celeb-1M (5.8 M, 85 k)	99.83

[129]	2019	CLMLE	CLMLE Loss CASIA WebFace (494 k, 10 k)	99.62
[130]	2019	PDSN	Pairwise Contrastive Loss CASIA WebFace (494 k, 10 k)	99.20
[131]	2020	MML	MML Loss VGGFace 2 (3.05 M, 8 K)	99.63
[132]	2020	IAM	IAM loss CASIA WebFace (494 k, 10 k)	99.12
[133]	2020	RCM loss	Rotation Consistent Margin loss CASIA WebFace (494 k, 10 k)	98.91
[134]	2020	ACNN	ArcFace Loss DeepGlint -MS1M (3.9 M, 86 K)	99.83
[135]	2020	LMC SDLMC DLMC	LMC loss SDLMC loss DLMC loss CASIA WebFace (494 k, 10 k)	98.13 99.03 99.07

Article OPEN ACCESS

REFERENCES

[1] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection.” arXiv, Jun. 18, 2020. Accessed: Aug. 26, 2023. [Online]. Available: http://arxiv.org/abs/2001.00179

[2] Z. Akhtar, D. Dasgupta, and B. Banerjee, “Face Authenticity: An Overview of Face Manipulation Generation, Detection and Recognition,” SSRN Journal, 2019, doi: 10.2139/ssrn.3419272.

[3] Y. Mirsky and W. Lee, “The Creation and Detection of Deepfakes: A Survey.” Sep. 13, 2020. doi: 10.1145/3425780.

[4] S. Market-Growth and F. Trends, “Mordor Intelligence,” URL: https://www. mordorintelligence. com/industry-reports/smartwatch-market [accessed 2020-01-28], 2020.

[5] E. Nowroozi, S. Seyedshoari, M. Mohammadi, and A. Jolfaei, “Impact of Media Forensics and Deepfake in Society,” 2022, pp. 387–410. doi: 10.1007/978-3-031-10706-1_18.

[6] M. Khichi and R. K. Yadav, “Analyzing the Methods for Detecting Deepfakes,” in 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), IEEE, 2021, pp. 340–345.

[7] G. Guo and N. Zhang, “A survey on DL based face recognition,” Computer vision and image understanding, vol. 189, p. 102805, 2019.

[8] A. Haliassos, K. Vougioukas, S. Petridis, and M. Pantic, “Lips don’t lie: A generalisable and robust approach to face forgery detection,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 5039–5049.

[9] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, “Past, present, and future of face recognition: A review,” Electronics, vol. 9, no. 8, p. 1188, 2020.

[10] “Sensors | Free Full-Text | Face Recognition Systems: A Survey.” Accessed: Oct. 11, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/20/2/342

[11] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa, “Disguised faces in the wild,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 1–9.

[12] B. Maze et al., “Iarpa janus benchmark-c: Face dataset and protocol,” in 2018 international conference on biometrics (ICB), IEEE, 2018, pp. 158–165.

[13] Y. Liu, A. Jourabloo, and X. Liu, “Learning deep models for face anti-spoofing: Binary or auxiliary supervision,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 389–398.

[14] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, “Unsupervised domain adaptation for face anti-spoofing,” IEEE Transactions on Information Forensics and Security, vol. 13, no. 7, pp. 1794–1809, 2018.

[15] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks”.

[16] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “Faceforensics++: Learning to detect manipulated facial images,” in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1–11.

[17] S. Zhang et al., “A dataset and benchmark for large-scale multi-modal face anti-spoofing,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 919–928.

[18] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, “Celeb-df: A large-scale challenging dataset for deepfake forensics,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3207–3216.

[19] O. Elharrouss, N. Almaadeed, and S. Al-Maadeed, “LFR face dataset: Left-Front-Right dataset for pose-invariant face recognition in the wild,” in 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), IEEE, 2020, pp. 124–130.

[20] 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, 2023.

[21] B. Huang et al., “Masked face recognition datasets and validation,” in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 1487–1491.

[22] L. Liu, H. Yu, S. Wang, L. Wan, and S. Han, “Learning shape and texture progression for young child face aging,” Signal Processing: Image Communication, vol. 93, p. 116127, 2021.

[23] P. K. Chandaliya and N. Nain, “ChildGAN: Face aging and rejuvenation to find missing children,” Pattern Recognition, vol. 129, p. 108761, 2022.

[24] “Khairnar, Smita, et al. "Face Liveness Detection... - Google Scholar.” Accessed: Aug. 26, 2023. [Online]. Available:https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Khairnar%2C+Smita%2C+et+al.+%22Face+Liveness+Detection+Using+Artificial+Intelligence+Techniques%3A+A+Systematic+Literature+Review+and+Future+Directions.%22+Big+Data+and+Cognitive+Computing+7.1+%282023%29%3A+37.&btnG=

[25] J. Hernandez-Ortega, J. Fierrez, A. Morales, and P. Tome, “Time analysis of pulse-based face anti-spoofing in visible and NIR,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 544–552.

[26] A. Kohli and A. Gupta, “Detecting deepfake, faceswap and face2face facial forgeries using frequency cnn,” Multimedia Tools and Applications, vol. 80, pp. 18461–18478, 2021.

[27] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, “CNN-generated images are surprisingly easy to spot... for now,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 8695–8704.

[28] J. Thies, M. Zollhöfer, and M. Nießner, “Deferred neural rendering: Image synthesis using neural textures,” Acm Transactions on Graphics (TOG), vol. 38, no. 4, pp. 1–12, 2019.

[29] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, “Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2889–2898.

[30] B. Dolhansky et al., “The DeepFake Detection Challenge (DFDC) Dataset.” arXiv, Oct. 27, 2020. Accessed: Nov. 12, 2023. [Online]. Available: http://arxiv.org/abs/2006.07397

[31] S. R. Arashloo, “Unseen face presentation attack detection using sparse multiple kernel Fisher null-space,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, pp. 4084–4095, 2020.

[32] “Snapshot.” Accessed: Aug. 27, 2023. [Online]. Available: https://www.kaggle.com/c/deepfake-detection-challenge

[33] K. Kotwal et al., “Domain-specific adaptation of CNN for detecting face presentation attacks in NIR,” IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 1, pp. 135–147, 2022.

[34] D. Wang et al., “Wild Face Anti-Spoofing Challenge 2023: Benchmark and Results.” arXiv, May 04, 2023. Accessed: Nov. 12, 2023. [Online]. Available: http://arxiv.org/abs/2304.05753

[35] S. Tariq, S. Lee, H. Kim, Y. Shin, and S. S. Woo, “Detecting both machine and human created fake face images in the wild,” in Proceedings of the 2nd international workshop on multimedia privacy and security, 2018, pp. 81–87.

[36] N.-T. Do, I.-S. Na, and S.-H. Kim, “Forensics face detection from GANs using convolutional neural network,” ISITC, vol. 2018, pp. 376–379, 2018.

[37] X. Xuan, B. Peng, W. Wang, and J. Dong, “On the generalization of GAN image forensics,” in Chinese conference on biometric recognition, Springer, 2019, pp. 134–141.

[38] P. Yang, R. Ni, and Y. Zhao, “Recapture image forensics based on Laplacian convolutional neural networks,” in Digital Forensics and Watermarking: 15th International Workshop, IWDW 2016, Beijing, China, September 17-19, 2016, Revised Selected Papers 15, Springer, 2017, pp. 119–128.

[39] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, “Learning self-consistency for deepfake detection,” in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 15023–15033.

[40] M. Wang and W. Deng, “Deep face recognition: A survey,” Neurocomputing, vol. 429, pp. 215–244, 2021.

[41] M. Mehdipour Ghazi and H. Kemal Ekenel, “A comprehensive analysis of deep learning based representation for face recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016, pp. 34–41.

[42] N. E. Khalifa, M. Loey, and S. Mirjalili, “A comprehensive survey of recent trends in deep learning for digital images augmentation,” Artificial Intelligence Review, pp. 1–27, 2022.

[43] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” Global Transitions Proceedings, vol. 3, no. 1, pp. 91–99, 2022.

[44] X. Wang, K. Wang, and S. Lian, “A survey on face data augmentation for the training of deep neural networks,” Neural computing and applications, vol. 32, no. 19, pp. 15503–15531, 2020.

[45] I. Masi, T. Hassner, A. T. Tran, and G. Medioni, “Rapid synthesis of massive face sets for improved face recognition,” in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), IEEE, 2017, pp. 604–611.

[46] I. Masi, A. T. Trần, T. Hassner, J. T. Leksut, and G. Medioni, “Do we really need to collect millions of faces for effective face recognition?,” in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14, Springer, 2016, pp. 579–596.

[47] Y. Qian, W. Deng, and J. Hu, “Task specific networks for identity and face variation,” in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 271–277.

[48] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, and J. Kim, “Rotating your face using multi-task deep neural network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 676–684.

[49] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, “Towards open-set identity preserving face synthesis,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6713–6722.

[50] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, “CVAE-GAN: fine-grained image generation through asymmetric training,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2745–2754.

[51] W. Chai, W. Deng, and H. Shen, “Cross-generating gan for facial identity preserving,” in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 130–134.

[52] Y. Shen, P. Luo, J. Yan, X. Wang, and X. Tang, “Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 821–830.

[53] J. Zhao et al., “Dual-agent gans for photorealistic and identity preserving profile face synthesis,” Advances in neural information processing systems, vol. 30, 2017.

[54] J. Yang, S. E. Reed, M.-H. Yang, and H. Lee, “Weakly-supervised Disentangling with Recurrent Transformations for 3D View Synthesis”.

[55] Z. Zhu, P. Luo, X. Wang, and X. Tang, “Deep Learning Identity-Preserving Face Space,” in 2013 IEEE International Conference on Computer Vision, Sydney, Australia: IEEE, Dec. 2013, pp. 113–120. doi: 10.1109/ICCV.2013.21.

[56] L. Tran, X. Yin, and X. Liu, “Disentangled Representation Learning GAN for Pose-Invariant Face Recognition,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE, Jul. 2017, pp. 1283–1292. doi: 10.1109/CVPR.2017.141.

[57] Z. Zhang, X. Chen, B. Wang, G. Hu, W. Zuo, and E. R. Hancock, “Face Frontalization Using an Appearance-Flow-Based Convolutional Neural Network,” IEEE Trans. on Image Process., vol. 28, no. 5, pp. 2187–2199, May 2019, doi: 10.1109/TIP.2018.2883554.

[58] “APA: Adaptive Pose Alignment for Pose-Invariant Face Recognition | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Aug. 27, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/8620989/?denied=

[59] W. Zhang, X. Zhao, J.-M. Morvan, and L. Chen, “Improving Shadow Suppression for Illumination Robust Face Recognition.” arXiv, Oct. 13, 2017. Accessed: Aug. 27, 2023. [Online]. Available: http://arxiv.org/abs/1710.05073

[60] O. Russakovsky et al., “Imagenet large scale visual recognition challenge,” International journal of computer vision, vol. 115, pp. 211–252, 2015.

[61] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, vol. 25, 2012.

[62] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[63] C. Szegedy et al., “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[64] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[65] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.

[66] Y. Rao, J. Lu, and J. Zhou, “Learning discriminative aggregation network for video-based face recognition and person re-identification,” International Journal of Computer Vision, vol. 127, pp. 701–718, 2019.

[67] E. Hand and R. Chellappa, “Attributes for improved attributes: A multi-task network utilizing implicit and explicit relationships for facial attribute classification,” in Proceedings of the AAAI conference on artificial intelligence, 2017.

[68] E. Xing, M. Jordan, S. J. Russell, and A. Ng, “Distance metric learning with application to clustering with side-information,” Advances in neural information processing systems, vol. 15, 2002.

[69] C. Whitelam et al., “Iarpa janus benchmark-b face dataset,” in proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 90–98.

[70] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 815–823.

[71] W. Liu, Y. Wen, Z. Yu, and M. Yang, “Large-margin softmax loss for convolutional neural networks,” arXiv preprint arXiv:1612.02295, 2016.

[72] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “Sphereface: Deep hypersphere embedding for face recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 212–220.

[73] W. Liu et al., “Deep hyperspherical learning,” Advances in neural information processing systems, vol. 30, 2017.

[74] V. Jain, “The Indian face database,” http://vis-www. cs. umass. edu/vidit/IndianFaceDatabase/, 2002.

[75] Y. Guo and L. Zhang, “One-shot face recognition by promoting underrepresented classes,” arXiv preprint arXiv:1707.05574, 2017.

[76] Z. Ding, Y. Guo, L. Zhang, and Y. Fu, “One-shot face recognition via generative learning,” in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 1–7.

[77] L. Wang, Y. Li, and S. Wang, "Feature learning for one-shot face recognition," in 2018 25th IEEE international conference on image processing (ICIP), IEEE, 2018, pp. 2386–2390.
[78] Y. Wu, H. Liu, and Y. Fu, "Low-shot face recognition with hybrid classifiers," in Proceedings of the IEEE international conference on computer vision workshops, 2017, pp. 1933–1939.
[79] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in 2018 IEEE international workshop on information forensics and security (WIFS), IEEE, 2018, pp. 1–7.
[80] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," Interfaces (GUI), vol. 3, no. 1, pp. 80–87, 2019.
[81] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking," in 2018 IEEE International workshop on information forensics and security (WIFS), IEEE, 2018, pp. 1–7.
[82] J. Donahue et al., "Long-term recurrent convolutional networks for visual recognition and description," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 2625–2634.
[83] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," arXiv preprint arXiv:1811.00656, 2018.
[84] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 8261–8265.
[85] J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, "Targeting ultimate accuracy: Face recognition via deep embedding," arXiv preprint arXiv:1506.07310, 2015.
[86] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 2892–2900.
[87] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," Advances in neural information processing systems, vol. 27, 2014.
[88] S. Gong, Y. Shi, N. D. Kalka, and A. K. Jain, "Video face recognition: Component-wise feature aggregation network (c-fan)," in 2019 International Conference on Biometrics (ICB), IEEE, 2019, pp. 1–8.
[89] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in BMVC 2015-Proceedings of the British Machine Vision Conference 2015, British Machine Vision Association, 2015.
[90] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1701–1708.
[91] J. Yang et al., "Neural aggregation network for video face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4362–4371.
[92] K. Kim, Z. Yang, I. Masi, R. Nevatia, and G. Medioni, "Face and body association for video-based face recognition," in 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018, pp. 39–48.
[93] Y. Rao, J. Lu, and J. Zhou, "Attention-aware deep reinforcement learning for video face recognition," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 3931–3940.
[94] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," ACM Transactions on intelligent systems and technology (TIST), vol. 7, no. 3, pp. 1–42, 2016.
[95] J. Grundström, "Face verification and open-set identification for real-time video applications," Master's Theses in Mathematical Sciences, 2015.
[96] X. Liu, B. V. K. Kumar, C. Yang, Q. Tang, and J. You, "Dependency-aware attention control for unconstrained face recognition with image sets," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 548–565.
[97] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), IEEE, 2018, pp. 1–6.
[98] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos," in 2019 IEEE 10th international conference on biometrics theory, applications and systems (BTAS), IEEE, 2019, pp. 1–8.
[99] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "Wilddeepfake: A challenging real-world dataset for deepfake detection," in Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 2382–2390.
[100] S. Fernandes et al., "Detecting deepfake videos using attribution-based confidence metric," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 308–309.
[101] U. A. Ciftci, I. Demir, and L. Yin, "Fakecatcher: Detection of synthetic portrait videos using biological signals," IEEE transactions on pattern analysis and machine intelligence, 2020.
[102] A. Gandhi and S. Jain, "Adversarial perturbations fool deepfake detectors," in 2020 international joint conference on neural networks (IJCNN), IEEE, 2020, pp. 1–8.
[103] L. Li et al., "Face x-ray for more general face forgery detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 5001–5010.
[104] X. Li et al., "Sharp multiple instance learning for deepfake video detection," in Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 1864–1872.
[105] M. S. Rana and A. H. Sung, "Deepfakestack: A deep ensemble-based learning technique for deepfake detection," in 2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom), IEEE, 2020, pp. 70–75.
[106] Z. Guo, L. Hu, M. Xia, and G. Yang, "Blind detection of glow-based facial forgery," Multimedia Tools and Applications, vol. 80, pp. 7687–7710, 2021.
[107] Z. Chen and H. Yang, "Attentive semantic exploring for manipulated face detection," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 1985–1989.
[108] A. Sedik et al., "An efficient cybersecurity framework for facial video forensics detection based on multimodal deep learning," Neural Computing and Applications, vol. 34, Jan. 2022, doi: 10.1007/s00521-021-06416-6.
[109] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1891–1898.
[110] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," arXiv preprint arXiv:1502.00873, 2015.
[111] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Web-scale training for face identification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 2746–2754.
[112] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, "Range loss for deep face recognition with long-tailed training data," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 5409–5418.
[113] B. Chen, W. Deng, and J. Du, "Noisy softmax: Improving the generalization ability of dcnn via postponing the early softmax saturation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5372–5381.
[114] H. Ben Fredj, S. Bouguezzi, and C. Souani, "Face recognition in unconstrained environment with CNN," The Visual Computer, vol. 37, pp. 217–226, 2021.
[115] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in Computer Vision–ECCV 2016: 14th

European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14, Springer, 2016, pp. 499–515.
[116] Y. Wu, H. Liu, J. Li, and Y. Fu, "Deep face recognition with center invariant loss," in Proceedings of the on Thematic Workshops of ACM Multimedia 2017, 2017, pp. 408–414.
[117] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Feature transfer learning for face recognition with under-represented data," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 5704–5713.
[118] R. Ranjan, C. D. Castillo, and R. Chellappa, "L2-constrained softmax loss for discriminative face verification," arXiv preprint arXiv:1703.09507, 2017.
[119] J. Deng, Y. Zhou, and S. Zafeiriou, "Marginal loss for deep face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 60–68.
[120] F. Wang, X. Xiang, J. Cheng, and A. L. Yuille, "Normface: L2 hypersphere embedding for face verification," in Proceedings of the 25th ACM international conference on Multimedia, 2017, pp. 1041–1049.
[121] Y. Liu, H. Li, and X. Wang, "Rethinking feature discrimination and polymerization for large-scale recognition," arXiv preprint arXiv:1710.00870, 2017.
[122 M. A. Hasnat, J. Bohné, J. Milgram, S. Gentric, and L. Chen, "von mises-fisher mixture model-based deep learning: Application to face verification," arXiv preprint arXiv:1706.04264, 2017.
[123] Y. Zheng, D. K. Pal, and M. Savvides, "Ring loss: Convex feature normalization for face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5089–5097.
[124] H. Wang et al., "Cosface: Large margin cosine loss for deep face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5265–5274.
[125] F. Wang, J. Cheng, W. Liu, and H. Liu, "Additive margin softmax for face verification," IEEE Signal Processing Letters, vol. 25, no. 7, pp. 926–930, 2018.
[126]X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," IEEE Transactions on Information Forensics and Security, vol. 13, no. 11, pp. 2884–2896, 2018.
[127] M. Hayat, S. Khan, S. W. Zamir, J. Shen, and L. Shao, "Gaussian affinity for max-margin class imbalanced learning," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6469–6479.
[128] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4690–4699.
[129] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Deep imbalanced learning for face recognition and attribute prediction," IEEE transactions on pattern analysis and machine intelligence, vol. 42, no. 11, pp. 2781–2794, 2019.
[130] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential siamese network," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 773–782.
[131] X. Wei, H. Wang, B. Scotney, and H. Wan, "Minimum margin loss for deep face recognition," Pattern Recognition, vol. 97, p. 107012, 2020.
[132] J. Sun, W. Yang, R. Gao, J.-H. Xue, and Q. Liao, "Inter-class angular margin loss for face recognition," Signal Processing: Image Communication, vol. 80, p. 115636, 2020.
[133] Y. Wu et al., "Rotation consistent margin loss for efficient low-bit face recognition," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 6866–6876.
[134] H. Ling, J. Wu, J. Huang, J. Chen, and P. Li, "Attention-based convolutional neural network for deep face recognition," Multimedia Tools and Applications, vol. 79, pp. 5595–5616, 2020.
[135] B. Wu and H. Wu, "Angular discriminative deep feature learning for face verification," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 2133–2137.
[136] A. Kale and O. Altun, "Face Age Synthesis: A Review on Datasets, Methods, and Open Research Areas," Pattern Recognition, p. 109791, 2023.
[137] "Khairnar, Smita, et al. "Face Liveness Detection... - Google Scholar." Accessed: Aug. 26, 2023. [Online]. Available:https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Khairnar%2C+Smita%2C+et+al.+%22Face+Liveness+Detection+Using+Artificial+Intelligence+Techniques%3A+A+Systematic+Literature+Review+and+Future+Directions.%22+Big+Data+and+Cognitive+Computing+7.1+%282023%29%3A+37.&btnG=
[138] R. Shao, P. Perera, P. C. Yuen, and V. M. Patel, "Federated generalized face presentation attack detection," IEEE Transactions on Neural Networks and Learning Systems, 2022.
[139] D. Zhang, P. An, and H. Zhang, "Application of robust face recognition in video surveillance systems," Optoelectronics Letters, vol. 14, no. 2, pp. 152–155, 2018.
[140] M. Taskiran, N. Kahraman, and C. E. Erdem, "Face recognition: Past, present and future (a review)," Digital Signal Processing, vol. 106, p. 102809, 2020.
[141] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4680–4689.
[142] C. Nagpal and S. R. Dubey, "A performance evaluation of convolutional neural networks for face anti spoofing," in 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.
[143] X. Yang et al., "Face anti-spoofing: Model matters, so does data," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3507–3516.
[144] J. Yang, A. Bulat, and G. Tzimiropoulos, "Fan-face: a simple orthogonal improvement to deep face recognition," in Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 12621–12628.
[145] S. Zhang, Z. Huang, D. P. Paudel, and L. Van Gool, "Facial emotion recognition with noisy multi-task annotations," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 21–31.
[146] "Full Text PDF." Accessed: Nov. 12, 2023. [Online]. Available: https://arxiv.org/pdf/2006.07397.pdf