Iraqi Journal for Computer Science and Mathematics

Volume 6 | Issue 1 Article 8

2025

Facial Swap Detection Based on Deep Learning: Comprehensive Analysis and Evaluation

Israa Mishkhal
School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia AND Department of Computer Science, Sciences College, University of Diyala, Iraq

Nibras Abdullah
Faculty of Computer Studies, Arab Open University, Saudi Arabia, n.faqera@arabou.edu.sa

Hassan h. Saleh
Department of Computer Science, Sciences College, University of Diyala, Iraq

Nur Intan Raihana Ruhaiyem
School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, 11800, Malaysia

Fadratul Hafinaz Hassan
School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia

Recommended Citation

Mishkhal, Israa; Abdullah, Nibras; Saleh, Hassan h.; Ruhaiyem, Nur Intan Raihana; and Hassan, Fadratul Hafinaz (2025) "Facial Swap Detection Based on Deep Learning: Comprehensive Analysis and Evaluation," *Iraqi Journal for Computer Science and Mathematics*: Vol. 6: Iss. 1, Article 8.
DOI: https://doi.org/10.52866/2788-7421.1229
Available at: https://ijcsm.researchcommons.org/ijcsm/vol6/iss1/8

IRAQI JOURNAL FOR COMPUTER SCIENCE AND MATHEMATICS 2025;6:109–123

REVIEW

Facial Swap Detection Based on Deep Learning: Comprehensive Analysis and Evaluation

Israa Mishkhal [a,b], Nibras Abdullah [c,*], Hassan h. Saleh [b], Nur Intan Raihana Ruhaiyem [a], Fadratul Hafinaz Hassan [a]

[a] School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, 11800, Malaysia
[b] Department of Computer Science, Sciences College, University of Diyala, Iraq
[c] Faculty of Computer Studies, Arab Open University, Saudi Arabia

ABSTRACT

In recent years, Advancements in Artificial Intelligence (AI), particularly deep learning (DL), have made great strides in the creation of highly realistic deepfakes, which manipulate facial forensics to generate convincing fake faces or expressions. These manipulations pose significant threats to individual privacy and the integrity of legal, political, and social institutions. In fact, several existing studies have recently pursued the development of machine learning techniques for detecting deepfake content, with the overarching aim of protecting the victim's privacy or curbing the rise of picture fabrication. Despite extensive research on DL-based deepfake detection systems, challenges such as detecting facial swaps under occlusion or subtle alteration remain insufficiently addressed. This study provides a comprehensive detailed evaluation of state-of the art DL approaches for detecting such manipulations, focusing on their strengths and limitations. Additionally, it reviews recent deepfake datasets (2019 to date) to identify their adequacy for adequacy for training and testing these models. By addressing the gaps and limitations in current existing methods and datasets, this study aims to pave the way for redefining DL-based detection techniques tailored to facial-swap-based deepfakes. It aspires to enhance the integrity and reliability of image media and contributes to the ongoing effort to mitigate the risk posed by advanced image manipulation.

Keywords: Deepfakes, Deep learning, Face manipulation, Face swap, Facial forensic, CNNs

1. Introduction

The manipulation of image and video content is not a new phenomenon. Recently, various editing tools, such as adobe lightroom, adobe Photoshop, have been widely utilized for various image modifications [1, 2]. However, these tools have faced significant limitations when it comes to realistically altering facial features in deepfake contexts. The processes are often complex, time-consuming processes, and require domain expertise. Recent advancements in deepfake techniques have revolutionized this landscape, significantly reducing the effort required for manipulation, especially facial swap [3]. Techniques such as FaceSwap (FS) and FaceApp (FA) are prime examples of these innovations [4]. These techniques are developed using artificial neural networks, relying on advanced DL approaches, such as Generative Adversarial Networks (GANs) and Autoencoder algorithms [5]. These methods enable the synthesis of manipulated faces in images and videos, producing convincingly realistic results [6]. While deepfake technologies offer opportunities for creative applications such as voice dubbing in videos [6] or digital try-ons for shopping [7], they also raise significant global security concerns. These technologies allow the manipulation of individuals' faces in media [6], leading to the potential for misuse in spreadingmisinformation, manipulating elections, or inciting social unrest [7].

Received 18 September 2024; revised 24 December 2024; accepted 31 December 2024.
Available online 13 February 2025

* Corresponding author.
E-mail address: n.faqera@arabou.edu.sa (N. Abdullah).

https://doi.org/10.52866/2788-7421.1229

As the capabilities of deepfake tools such as DeepFaceLab [8, 9], Zao [10], and FakeApp [11] have become more accessible, even non-professional users can now create highly realistic face swaps. This has heightened concerns about the harmful consequences that often outweigh their potential positive uses. Manipulation techniques can be classified into four main types [12]. The first, **Face Generation**, involves creating entirely new facial images or videos that do not correspond to any real individuals. The second, **Facial Attribute changes**, refers to modifications in features such as age, gender, hair color, glasses, or the use of face masks. The third, **Facial Swap**, entails replacing one face with another, blending features and expressions to create a seamless, realistic deepfake. The final category, **Expression Swap**, transfers the facial expression of one person to another, altering their displayed emotions to match those of the original individual.

Deep learning (DL) has a significant impact on both the generating and detection of deepfakes, including facial swaps [6]. DL algorithms analyze facial features, movements, and expressions to manipulate target faces, making deepfake generation more efficient. However, the same robustness that makes DL techniques effective for generating realistic deepfakes also makes detection more challenging [13, 14]. The increasing sophistication of these technologies underscores the need for advanced detection systems capable of identifying deepfakes in real-world, unconstrained environments [10]. In recent years, facial swap detection has emerged a critical area of research aimed at countering malicious applications. A wide variety of techniques have been proposed, many of which tread facial swap detection as a binary classification problem and focus on designing effective features [15–20]. Early methods leveraged artifacts [15], frequency analysis [16], symmetry [81], and local differences [21]. More advanced approaches, such as EfficientNet [17] and Xception [22], have also been applied to the task of detecting facial swap. While existing research has provided important insights, several challenges remain in ensuring accurate and reliable detection of deepfake facial swaps. Currently methods often struggle to detect high-quality, realistic deepfakes, particularly in environments with uncontrolled variables. This highlights the need for improved detection systems and more comprehensive datasets for training robust deep learning models.

This paper aims to contribute to this field in several ways:

1. Providing a comprehensive overview of facial swap technology, emphasizing the importance of generating and detecting facial swaps using DL approaches.
2. Assessing and contrasting various DL techniques based on their structures and performance evaluation metrics.
3. Analyzing existing datasets in this field, examining their diversity, size, and quality, and highlighting their implications for training more effective DL models.
4. Exploring new directions and future challenges in facial swap detection systems, highlighting areas that require further development.

The remainder of the paper is organized as follows: Section 2 reviews related works focused on detecting and generating facial manipulations, particularly facial swaps. Section 3 provides a list of datasets commonly used to evaluate most facial swaps detection approaches. Section 4 explores different DL-based detection methods for facial swaps. Section 5 discusses the weakness and limitation of current detection approaches and outlines potential future directions. Finally, Section 6 presents the conclusions of the paper.

2. Overview of deepfake technology

Deepfake has gained widespread attention around the world due to videos and images that have become easily getting on the internet for visual manipulation, especially facial swap [23]. Leveraging DL in the deepfake generation and detection has revolutionized the field of image and video synthesis (see Fig. 1) [24].

2.1. Deepfake generation

Deepfake generation, particularly facial swap, typically involves three stages [25]. Firstly, tools to detect a target face and compare to get a similar face appearance in the facial dataset. Secondly, using methods to replace facial details such as eyes, nose, and mouth and further adjusts the lighting and color of the candidate facial to match the appearance of input images. Finally, the generator deepfake system measures the distance between two selected faces in the images in able to replace them. Table 1 shows the most facial swap generation excited approaches from 2019 until 2023. It refers to the automated process of replacing the face of the target image with another person (as shown in Fig. 2).

Due to DL approaches having achieved realistic results; these become popular for generator deepfakes. They showed DL can be applied with automated digital manipulation. Many applications found to swap

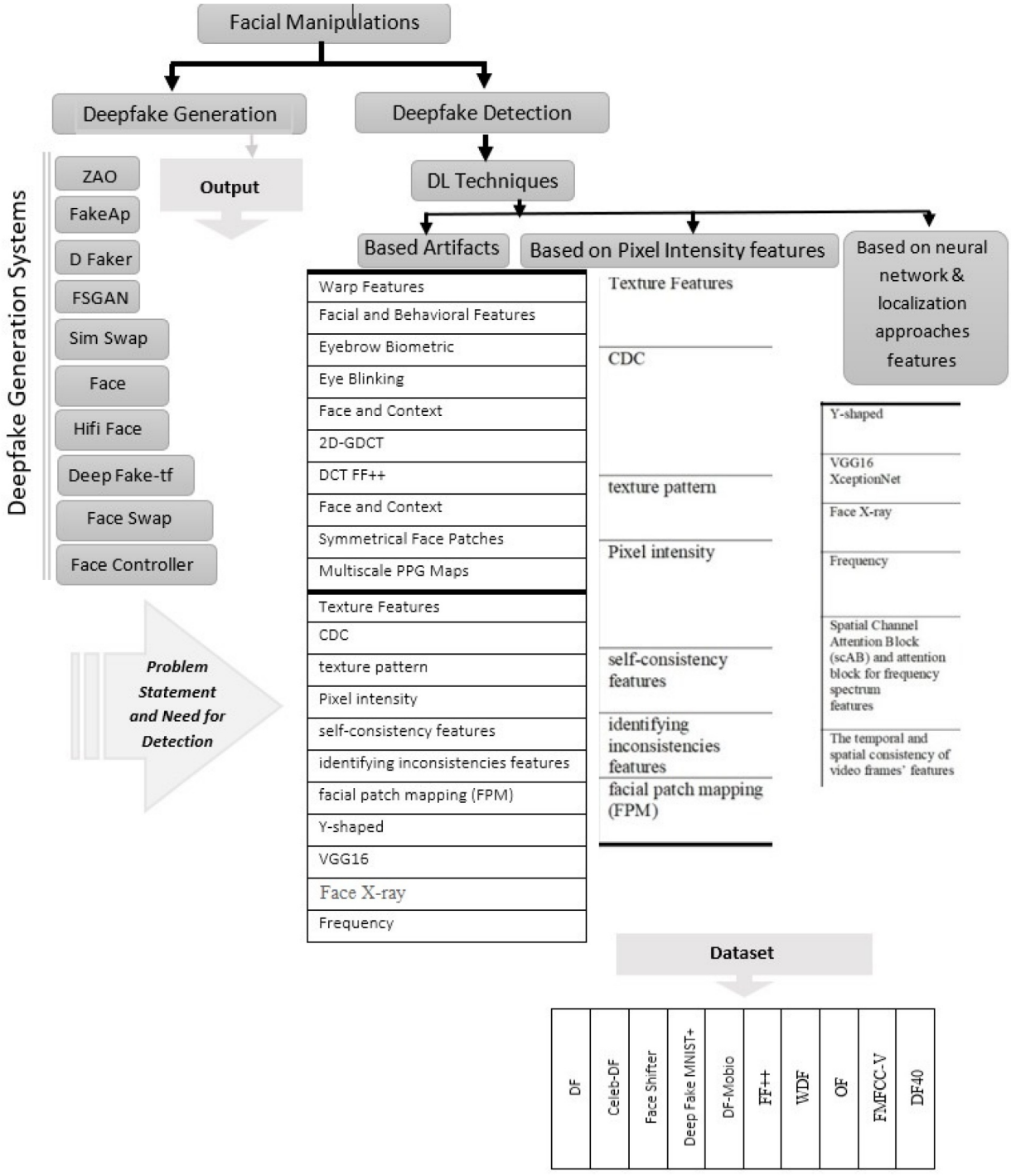

Fig. 1. Overview of facial manipulation.

faces, such as FakeApp [11] and Faceswap [27], that have presented to produce easy and quick deepfakes with more persuasive results by face swap in a video or image. They presented techniques typically utilizing Encoder-Decoder approaches. To extract features of the face, usually Encoder-Decoder has utilized to reconstruct faces. Facial swap needs to use two pairs of encoder and decoder: the encoder approach is to train the features in the target facies, and the decoder is utilized to swap selected faces by utilizing to regenerate the target faces with the features of the source face. The result keeps the facial expressions of the target faces, and near to be realistic [26]. Other facial swap application that are more popular used

Table 1. The most popular deepfake generation systems.

Application name	Ref	Features	Approach	Dataset Size	limitations
FSGAN 2019	[31]	VGG-19 CNNs	GANs	128 × 128 256 × 256	the approach has struggled to handle partial occlusion & expression cases
Faceswap-2020	[27]	MTCNN for align faces for fine-grained permutation	Autoencoders-GANs	854 × 480	It has struggled to handle faces in real-time, high-quality, and occlusion
ZAO-2020	[10]	VGGface	GANs	256 × 256	it has struggled when dealing with a complex video or rapid movements
DFaker-2020	[28]	VGGface	GAN	64*64	It has artifacts when using low quality / complex scenes.
Faceswap-GAN-2020	[30]	VGGFace & MTCNN	GAN	256 × 256	It left some artifacts and blending around the boundaries of the facial swap.
FaceShifter-2020	[14]	Attributes (face, occlusions, lighting or styles)	GAN	256*256	Stripped artifact
SimSwap-2020	[34]	simple feed-forward face generation network	3DMM	224*224	Black holes in a big-size facial swap image & the direction of the eye gaze does not match the target image
FaceController-2021	[32]	VGGface	3DMM	224 × 224	Struggle with occluded faces or partially visible
HifiFace-2021	[33]	3D Shape-Aware Identity Extractor & SFF	3DMM	512*512	Difficult to handle the occlusion
DeepFake-tf-2024	[29]	FCN-8s-VGGarchitecture	(FCN)	64 × 64-128 × 128 256 × 256	In complex videos, the result shows artifacts & blurriness. also, the approach has struggled to handle Occlusion cases
FaceSwapGAN-2024	[30]	VGGFace	GANs	256 × 256	Occlusion handling & artifacts

Fig. 2. Facial swap based on a DL approach [26].

because to generate a realistic facial swap such as ZAO [10], DFaker [28], DeepFaceLab [8], DeepFake-tf [29] and FaceSwapGAN [30]. These applications utilize to replace any person's face with another without extensive knowledge of the technology, leading to the creation of numerous deepfakes.

The most proposed generation deepfake approaches have achieved good performances using DL, however they have some limitations. First, expressions lost when these approaches are replacing the input face with the target. Also, the replaced faces look unnatural due to needing to match the faces details to generate good results [24].

Several studies focused on advances in facial swap using DL approaches. In 2019, Nirkin et al introduced as a facial swap generation based on GANs models, called FSGAN. This approach can swap faces without requiring dataset to train [31]. Also, in 2020 introduced a new study for generation deepfake based on GAN which adds adversarial and perceptual losses to

VGGface. It made the direction of the eyeball to be more realistic and try to remove artifacts to product a high-quality output deep fake [30]. Moreover, two existing facial swap approaches utilized 3DMM such as FaceController and HifiFace. The first one appeared a block holes in a big size image when swapping faces [32]. The other cannot handle occlusion cases [33].

In occlusion cases, facial swap is always challenging. Occlusions in the facial region mean the face covered with glasses, hair, a hat, hand or anything else. So, it results in facial swap with visual artifacts and inconsistencies. In 2020, the paper [14] proposed a new approach that reduced artifacts to appear the generation deep fakes as realistic. Additionally, Chen et al presented a new approach based the ID Injection Module (IIM) of facial swap. However, the quality of facial swap has been effaced by the quality of the datasets [34]. In 2023, Rosberg et al. presented a new technique to swap faces under occlusions based on a conditional GAN model, and it achieved a highly realistic result when compared with the previous methods [35].

2.2. Deepfake detection systems

Deepfake detection systems have emerged as a significance research area in recent years to tackle the potentially malicious system. Several models have been proposed for detecting facial swap using deep learning (DL) techniques. Facial swap detection systems based on DL approaches have gained significant attention in recent years [13]. Li et al. proposed the CNN models such as ResNet152, ResNet101, ResNet50, and VGG16 for detecting manipulations of facial landmarks in videos [15]. However, these systems exhibit reduced accuracy when tested on highly compressed videos. In 2021, Zendran and Rusiecki introduced a novel CNN-based method for features extraction at the frame level, combined with RNN methodologies for short-duration video [13]. The paper [36] presented an encoder –decoder network, FAB-Net, which combines facial and behavioral biometrics using VGG-16 for computation. Other papers [22, 37] have also employed CNN based method for facial swap detection.

Despite these advancements, there remains a need to improve detection accuracy further. However, these proposed systems exhibit decreased detection accuracy when tested them on multi-time compressed videos. Another approach presented in [22] introduced a novel CNN based method for extracting features at the frame level and utilized RNN methodology to detect facial swap. However, this technique is primarily applicable to short duration videos. Montserrat et al. proposed a multi-task CNN for detection facial swap in video, incorporating the use of Automatic Face Weighting (AFW) to discard false detected facial [38]. Nevertheless, this approach does not account for predications from multiple video frames. Agarwal et al. [36] proposed an encoder–decoder network called FAB-Net, which combines facial and behavioral biometrics to detect facial swap. VGG-16 is used for facial features computation in this approach. Other papers [22] and [37] have also employed CNN based method for facial swap detection, but there is still a need to improve accuracy in these approaches.

3. Deep learning techniques for facial swap detection

The task of facial swap detection necessitates the examination of their content to decide whether they have undergone alterations or remain in their original state. The facial swap (deepfake) is commonly approached as a binary classification problem namely: real or manipulated faces [72]. Therefore, researchers leverage the identification of extracted distinctive features that differentiate between authentic and manipulated content. So, they employ various approaches to extract these features such as DL algorithms, to identify inconsistencies and artifacts present in deepfakes. DL approaches have improved efficacy in automatically extracting features and their capability to operate at high speed [73]. The detection process typically extracts information of facial individual, followed by the utilization of a selected DL approach to find relevant features [74]. Subsequently, the extracted features are employed for detecting as genuine or fake.

In this section, we explore facial swap detection approaches based on DL, classifiers, noteworthy results, and assessed datasets [75]. Numerous features are utilized for the detection of manipulated faces using DL approaches. As shown in Table 2, the majority of deepfake detection methodologies rely on artifact features that are left by (GANs) during the generation process of deepfakes. These features serve as compelling evidence of manipulations and can be extracted for detection approaches [76]. These artifacts can be categorized into biometric and biological, and visual irregularities features (as shown in Fig. 3). In 2019, Yiru Zhao et al. proposed another approach that utilizes warp features to detect manipulated facial swap. It uses the ResNet-50 CNN architecture for detection [15]. The ResNet101 and VGG-CNNs architectures based on facial and behavioral features were introduced in 2020 [36]. This

Table 2. The most popular DL approaches for facial swap detection based on artifact features.

FE approach	Ref.	Year	method	Classifier	Dataset	Acc.	weakness
artifacts	[15]	2019	Warp Features	ResNet-50	DF-TIMIT	0.99	Rely on resolution inconsistency in face warping
	[36]	2020	Facial and Behavioral Features	ResNet101 VGG	Celeb-DF FF DFD DFDC-P	0.99 0.99 0.93 0.95	Relay on high-quality references videos
	[77]	2020	Eyebrow Biometric	LightCNN, ResNet, DenseNet, SqueezeNet	Celeb-DF	0.88	Relay on eyebrow matching. It is not the best choice for long term
	[78]	2020	Eye Blinking	DeepVision	FF	0.87	It limited by mental illnesses or problems.
	[79]	2020	Face and Context	XceptionNet	FF-DF Celeb-DF-v2	0.997 0.66	low contrast or blurry features
	[51]	2021	2D-GDCT	FCN	FF + + Celeb-DF	0.76 0.75	It has limitation on detecting the neural texture facial manipulation
	[52]	2021	DCT FF + +	SPSL(Xception)	FF + + Celeb-DF	0.969 0.768	Relay on combining spatial image and phase spectrum to capture the up-sampling artifacts of face
	[80]	2021	Face and Context	ResNet-18	DF F2F	0.971 0.974	It shows a decrease in accuracy as the video compression rate increases
	[81]	2021	Symmetrical Face Patches	DRN	FF + + DF-TIMIT DFD DFDC Celeb-DF	0.99 0.95 0.85 0.552 0.625	It has limitation on detecting the low quality data
	[53]	**2023**	Multiscale PPG Maps	EfficientNetV2	FF + +	0.90	Rely on the subtle changes in skin color caused by cardiac activity.

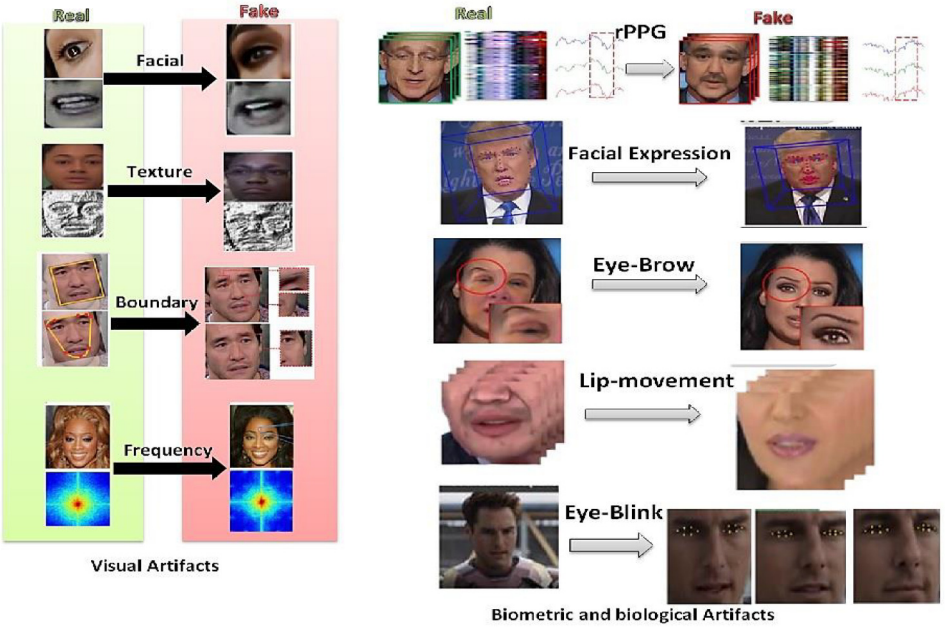

Fig. 3. Artifact features [41].

approach relies on high-quality reference videos, which can belimitation. Hoang Nguyen and Reza Derakhshani proposed LightCNN, ResNet, DenseNet, and SqueezeNet architectures to detect facial swap, but it may not be the best choice for long-term applications [77]. Other proposed approach that relayed on analyzing eye blinking called Deep-Vision approach. It has a limitation on mental illnesses or health problems [78]. The XceptionNet architecture is utilized facial and context features to detect facial swap. It achieved the highest accuracy when applied on FF-DF dataset, but it has limitations in dealing with low contrast or blurry features [79]. Aditi Kohli and Abhinav Gupta presented a new DL approach called FCN in 2021, which is based on Frequency and uses the 2D-GDCT extraction method. This approach achieved low accuracy when applied on FF + + and Celeb-DF datasets due to limitations in detecting neural facial manipulation [51]. The Spatial-Phase Shallow Learning (SPSL) based on the Xception-CNN architecture is introduced using DCT FF + + method. However, this approach has a limitation due to capturing the up-sampling artifacts of the face [52]. Kim and Cho introduced the Resnet-18 CNN architecture, which is based on analyzing the face and its context [80]. It achieved high performances when applied to the DF and F2F datasets. However, when it applies this approach to video compression, it shows a decrease in accuracy. Other authors presented another technique based on analyzing symmetrical face patches using a Dilated Residual Network (DRN) [81]. It is tested on several datasets and achieves the highest accuracy on FF + +. However, it has a limitation in its performance when detecting low-quality datasets. In 2023, Jiahui Wu et al. presented a new technique that utilizes EfficientNetV2 architecture [53]. This technique is based on Multiscale PPG and achieves 90% by relying on subtle changes in skin color caused by cardia activity.

The other detection approaches have relied on pixel intensity (see Table 3). Falko Matern et al. utilized a multi-Layer perceptron (MLP) and logistic regression (LogReg) based on texture features extraction. This approach is applying to images that meet certain prerequisites, such as having open eyes and visible teeth [39]. Yang et al. proposed a MTD network for deepfake detection. The proposed system is utilized to extract and integrate different information of multi scale texture using Center Difference Convolution technique (CDC). It achieved a high proficiency on high quality datasets, such as Celeb-DF, DeeperForensics –1.0, and DFDC, but it has some limitations such as low-quality datasets and the lack of interpretability [44]. In 2022, a new technique was proposed based on texture patterns that analyze texture inconsistencies using a CNN-based model called LBPNet [40]. However, this approach relies on analyzing texture inconsistency from a single frame, which can be a weakness. Wang et al. present a new approach called Facial Region Feature Descriptor (FFR-FD) based on seen features in facial images. This approach has limitations with low-resolution images or occlusions faces [82].

The other deepfake detection approaches based on neural networks to extract features and organize data. In 2021, Zhao et al. proposed a new approach based on CNN to extract the self-consistency features of original images, and then compare with the manipulated images. This approach has straggled on fake images that have the same consistent source properties in the entire manipulated images [83]. In 2022, Gangulay et al. introduced an approach depended on DL to improve visual attention technique for detecting manipulated images. This approach built using Xception network and focused on identifying inconsistencies characterizes in deepfake images. However, it has challenges detection on occlusions features such as closed eyes [84]. At the same year, YU et al. presented an alternative approach to train CNN utilizing Facial Patch Mapping (FPM). This technique divides an entire face to smaller regions that used to train Res-Net50 (Inception-v3). It trains five patches-based detectors. However, the struggle in this proposed system is hidden manipulations of specific regions of patches [85]. he majority of studies are binary classification deepfake detection approaches. However, there are several localization approaches (see Table 4). These can detect not only the real or manipulated facial swap but also identify the exact manipulated regions in images [41] (see Fig. 4).

In 2019, Nguyen et.al proposed a new of multi-task learning approach to detect and exact manipulated regions in facial images using a Y-shaped decoder based on CNN. This decoder is used to share information between multi-tasks. However, the performance of this approach was insignificant due to the training data [86]. In 2020, Dang et.al also used CNN to build their approach for detecting and localization deepfake regions in facial [87]. Li et al. introduced an approach using the face X-ray to detect considering noise and error levels in boundaries artifacts. This approach utilized FCN to detect the boundaries of deepfake images. However, the performance of this approach has struggle with low-quality dataset [88]. In 2022, Wang et al proposed a new novel to detect deepfakes and manipulated localization based on

Table 3. The most popular DL approaches for facial swap detection based on pixel intensity features.

FE approach	Ref.	Year	method	Classifier	Dataset	Acc.	weakness
Pixel intensity	[39]	2019	Texture Features	MLP, LogReg	ProGAN FF + +	0.85 0.86	applicable to images meeting certain prerequisites (e.g. open eyes, visible teeth)
	[44]	2021	CDC	MTD network	DFDC FF + + (C23,C40) Celeb-DF DeeperForensics-1.0	0.99	Rely on multi-scale texture difference information
	[83]		self-consistency features	ResNet-34	DF, F2F, FS, NT, FF + +	0.99	Manipulated images keep the same consistent source properties.
	[40]	2022	texture pattern	CNN-Based model (LBPNet)	FF + + Celeb-DF DFDC-P	0.99 0.92 0.80	Rely on analyzing texture inconsistency from a single frame.
	[82]	2022	Pixel intensity	FFR-FD	DF FF + + DFD CeLeb-DF DFDC	0.99 0.92 0.85 0.82 0.88	DF, FF + +, DFD achieved the accuracies using FAST&BRIEF, and other using SURF. In general, this approach has restrictions using low-quality or occlusions.
	[84]		identifying in-consistencies features	Xception net	FF + + CeLeb-DF	0.70	occlusions such as closed eyes
	[85]		facial patch mapping (FPM)	Res-Net50 (Inception-v3)	TIMIT (LQ,HQ) Celeb-DF	0.98	hidden manipulations of a specific regions of patches

Table 4. The most popular DL approaches for facial swap detection based on neural network & localization approaches features.

FE approach	Ref.	Year	Method	Classifier	Dataset	Acc.	weakness
localization approaches	[86]	2019	Y-shaped	CNN	FF FF + +	0.76	Small training dataset
	[87]	2020	VGG16 XceptionNet	CNN	DFFD UADFV Celeb-DF	0.99 0.84 0.64	Achieved the good performance on their dataset
	[88]		Face X-ray	FCN	DFD DFDC Celeb-DF	0.95 0.80 0.80	Achieved AUC when training it on FF + + & B1
	[89]	2022	Frequency	M2TR	FF + + (HQ) FF + + (LQ) Celeb-DF SRDF	0.97 0.93 0.99 0.91	Relay on HQ datasets
	[41]	2023	Spatial Channel Attention Block (scAB) and attention block for frequency spectrum features	MTCNN	DF (HQ) DF (LQ) FS (HQ) FS (LQ) FF + + Celeb-DF DFDC-P	0.99 0.96 0.97 0.93 0.97 0.68 0.79	Manipulation locations have limited fake pixels (unseen).
	[47]	2024	The temporal and spatial consistency of video frames' features	Swin-Fake	Celeb-DF FaceShifter DFDC	0.89 0.93 0.72	Rely on the spatial consistency that can be weakness when dealing varied datasets

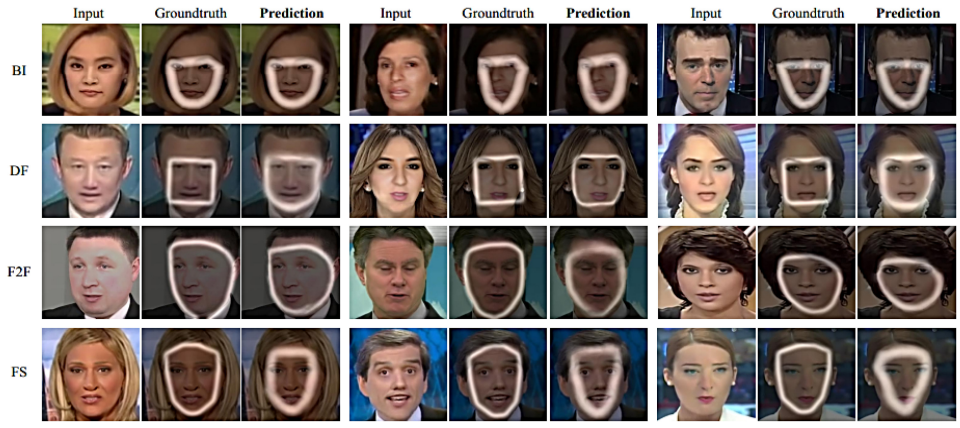

Fig. 4. Detect localization manipulated region [88].

frequency. Multimodal Multiscale Transformer (M2TF) employed two-stream architecture to capture different regions in images and filter out forged features. This method achieved a good performance on HQ dataset [89]. In 2023, Waseem et al introduced a new encoder-decoder approach to detect a manipulation localization utilizing a multi-attention mechanism. This approach used to capture frequency-related patterns in images with varying compression degradation. However, it has limitations in applicability when images do not have information about manipulation regions in the facial swap [41].

4. Datasets for facial swap detection

Facial swap, involving either replacing the face in a video or image with another person or changing motions from one target person to another, are currently the most prominent area of deep fake research [15, 16]. Recently, with the increasing apprehension surrounding the potential risks of misusing deepfake techniques, many research groups have made significant contributions by creating or collecting datasets to facilitate manipulation detection. In this section, we present a comprehensive overview of established datasets, especially focusing on facial swap. They can be categorized into two types depending on data size and variety visual fidelity and manipulation techniques. Table 5 provides a summary of the key characteristics of each public dataset categorized according to their types. Notably, these datasets encompass both real and deepfake, reflecting the nature of these particular types of manipulation.

4.1. First type based on data size

4.1.1. The deep fake detection challenge preview (DFDC-P) [39]

It highlighted a collaborative effort by three major companies (Facebook, Microsoft, and AWS) to advance deepfake detection methods in 2019 (see Fig. 5). This dataset comprises 1.131 authentic videos

Table 5. Summarize publicly available dataset of facial swap.

Dataset	Source	Deepfake	Real	General techniques
Based on Data Size				
DFDC-P [39]	Volunteer Actors	4.113	1.131	Four
Celeb-DF [43]	YouTube	5,639	590	One
FaceShifter [14]	YouTube	10,000	–	Two
DeepFakeMNIST + [48]	Internet	10,000	10,000	Two
DF-Mobio [50]	Internet	15,000	31,000	One
Based on Variety Visual Fidelity and Manipulation Techniques				
FF + + [35]	YouTube	4000	1000	Two
WDF [21]	Internet	3,509	3,805	One
KoDF [58]	Self-Recording	157,776	62,166	Six
OF [61]	Google Open Image	70,325	45,473	One
FMFCC-V [65]	Volunteer Actors	38,102	44,290	Four
DF-Platter [66]	YouTube	132,496	764	Three
DF40 [67]	–	300,000	100,000	Forty

Fig. 5. DFDC-P dataset [39].

and 4.113 manipulated videos created using facial-swap techniques.

This dataset incorporates three types of augmentations as perturbation methods. However, the specific synthesis algorithm employed in generating the dataset has not been disclosed. Several studies have utilized this dataset to test their approaches, including [36–42]. These studies reported varying accuracies, likely due to varies cases in this dataset, such as lighting condition, gender, face angles, and age.

4.1.2. Celeb-DF dataset [43]

This dataset collected of YouTube video featuring 59 celebrities, presented in 2020. It contains 590 authentic and 5,639 manipulated videos created using advanced DL technique defined as auto-encoder. The dataset encompasses a wide range of camera orientations, lighting conditions, and backgrounds. It is high quality dataset with no visible defects. Each video has duration of approximately 13 seconds and adheres to a standard frame rate of 30 frames per second.

Several studies have utilized this dataset to train or evaluate their proposed approaches. For example, [44] tested their method on Celeb-DF and achieved 99%, leveraging multi-scale texture information. The study [40] achieved 92% accuracy by focusing on texture inconsistency in single frames. The [45] reported the highest accuracy when using the dataset for training and testing, attributing this to the dataset's lack of diversity and its possession of super-set characteristics compared to others. In 2024, Solaiman, et al. proposed an approach based on CNN, achieved 75.07% accuracy on both realistic and manipulated videos due to enhancements in the dataset and a reduction in artifacts [46].

Fig. 6. FaceShifter dataset [14].

4.1.3. FaceShifter dataset [14]

It was generated using AI techniques on FF++ dataset in 2020. It consists of 10,000 meticulously deepfake videos of high quality, created using FaceShifter algorithm. This dataset (see Fig. 6) provides videos in three distinct compression qualities: Raw (C0, C23, C40). In 2023, a study proposed an approach that combined Residual U-Net with SCAB to detect deepfakes, which was evaluated on FaceShifter dataset [41]. In 2024, this dataset was utilized to evaluate the Swin-Fake approach, achieving a 98% accuracy rate by focusing on the boundary area where manipulation typically occur [47].

4.1.4. DeepFake MNIST+ dataset [48]

It was introduced by Huang et al in 2021. This dataset contains 10,000 videos featuring human faces displaying distinct expressions, and 10,000 real videos. These manipulated videos utilized the First Order motion Model (FOMM) [49] with various actions such as yaw, open mouth, and encompassing blink. Due to this dataset created using two public liveness detection application programming interfaces (APIs), most recently detection systems that failed to accurately detect were selectively included.

4.1.5. DF-Mobio dataset [50]

It is an extensive dataset comprising more than 46,000 videos in 2021. It includes 31,000 real videos and 15,000 deep fakes. This dataset consists of individuals engaging in direct communication using smartphones or laptops. These videos simulate virtual meeting conducted on platforms such as skype. GAN utilized to generate deep fakes; approximately 2,000 facial images were captured pre-person in the videos. The capturing process occurred at a frame rate of 8 frames pre-second. The input size was 256 × 256 pixels. This dataset was used to evaluate their approaches, and their experiments typically achieved results affected by unseen manipulations. However, their findings demonstrated

Fig. 7. WDF dataset [21].

improved detection accuracy when trained for attribution rather than using the typical binary method.

4.2. Second type based on variety visual fidelity and manipulation techniques

4.2.1. FaceForensics++ (FF++) datasets [22]

It introduced by Rössler et al. in 2019. It aims to facilitate research and development in the domain of deepfake detection of manipulated facial videos. It consists of over 4,000 manipulated videos created using various deepfake generation techniques, including both face swapping and facial reenactment methods. One notable aspect of the dataset is the inclusion of four different manipulation methods: FaceSwap, Deepfakes, Neural Textures, and Face2Face. Due to variety of manipulation techniques, several studies have utilized it to evaluate their approaches, including those by [39–53]. These studies have highlighted the variety of dataset's manipulations in benchmarking deepfake detection systems by providing different scenarios. Many researchers consider this dataset is an essential resource for evaluating and testing the effectiveness of various deepfake detection systems across a wide range of conditions, addressing both subtle and overt artifacts.

4.2.2. WildDeepfake (WDF) dataset [21]

It presents in 2020 and comprises about 7,314 facial swap videos: 3,805 real and 3,509 deep fake (see Fig. 7). Notably, these videos may feature more than ten individuals in each scene, rendering it an exceptional and invaluable resource for enhancing deep fake detection in real-world scenarios. Several studied have utilized the WDF dataset to evaluate their approaches, such as [22–57]. However, the performance of most of these approaches is affected in scenarios involving pixilate and blur.

4.2.3. Korean deepfake (KoDF) dataset [58]

It specifically features Korean actors and was introduced in 2021. This dataset includes a total of 62,166 real videos and 175,776 deepfakes generated using six different manipulation techniques. Three of these techniques are facial swapping methods: FSGAN, DeepFaceLab, and Faceswap, while the remaining techniques focus on facial reenactment. Numerous studied have widely used the KoDF dataset to evaluate and benchmark deepfake detection approaches due to its diverse range of manipulation techniques and size. For instance, [49–60] have utilized this dataset to assess the performance of their approaches due to its high-quality videos and incorporate both facial swapping and reenactment techniques.

4.2.4. OpenForensics (OF) dataset [61]

It is a substantial image dataset characterized by its diverse backgrounds. Real images were sourced from Google Open Images, while fake faces were generated using GAN-based. face techniques, Poisson blending, and color-matching algorithms in 2021. The primary aim OF is to provide high-resolution face images that exhibit enhanced visual quality and a more natural appearance. It comprises a total of 115,325 images, including 70,325 fakes and 45,473 real images. To simulate real-world challenges and create a more demanding test subset, various perturbations were applied to the dataset. These perturbations were categorized into several types, such as color manipulation, edge manipulation, block-wise distortion, image aliasing, convolution mask transformation, and external effects. These modifications ensure that the dataset is suitable for evaluating the robustness of deepfake detection approaches under diverse conditions. The OF has been extensively utilized in numerous studies to benchmark and validate detection approaches, such as [62–65]. Due to high-resolution and natural-looking fake images, the OF has played an essential role in promoting the development of more accurate and robust detection approaches.

4.2.5. The FMFCC-V dataset [66]

This dataset is Fake Media Forensics Challenge of China Society of Image and Graphics (FMFCC-V) that presents in 2022. FMFCC-V contains videos with diverse characteristics, including head poses, facial expressions, resolutions, backgrounds, and frame rates. The data's resolutions were 480, 720, and 1080 pixels and frame rates of 25 and 30 fps. This dataset includes 82,392 total videos, about 38,102 deepfake videos and 44,290 real. For the deepfake video utilized four various techniques namely: Faceswap_GAN, DeepFaceLab, Faceswap, and Recycle-GAN. This dataset

provides two versions of deepfake data defiantly: a long version with about 16 minutes of unperturbed videos and short version with 10 second videos, half of which have perturbation applied. It serves as a valuable tool for researchers and practitioners working on deepfake detection.

4.2.6. The DF-Platter dataset [67]

It contains total videos 133,260 that collect of YouTube in 2023. It comprises to 132,496 deep fakes and 764 real. The data encompass a board spectrum of expression, poses, backgrounds, lighting conditions, and occlusion. To generate deepfake videos, this dataset utilized three AI techniques namely: FaceShifter, FSGAN, and FaceSwap in high-resolution (HR) a low-resolution (LR) deepfakes. DF-Platter offers a diverse collection of videos featuring various subjects, characteristics, and deepfake techniques.

4.2.7. The DF40 dataset [68]

It was generated in 2024 using 40 distinct deepfake forgery techniques, including face swapping, reenactment, and full image synthesis. The dataset employs various generative methods such as HeyGen, DDIM, DiT, MidJourney6, PixArt-α, Stable Diffusion v2.1, SiT, StyleGAN2, StyleGAN3, StyleGAN-XL, VQGAN, Whichisreal, CollabDiff, e4e, StarGAN2, and StyleCLIP, and diffusion based models, resulting in a realistic, high resolution, and complex collection of videos. Additionally, DF40 includes videos with varying compression levels to simulate real-world scenarios, making it a valuable resource for researchers. This dataset supports the evaluation of deepfake detection approaches, especially those designed to identify advanced AI-driven manipulation techniques and address diverse data conditions. Several studies, such as [69–90] and [71] have utilized to evaluate their approaches.

5. Discussion and future directions

As shown in the previous section, most existing studies leverage DL approaches that have achieved high detection accuracies. The main reason for this success is the presence of artifacts or fingerprint information typical in facial swap generation systems. However, datasets with more realistic facial swaps, such as FaceShifter [14], SimSwap [34], FaceController [32], and FaceDancer [35], have not been used for training or evaluation, nor have detection systems achieved satisfactory performance on these datasets. Some of these datasets also contain facial swap under partial occlusions, resulting in high quality swaps. A new detection approach should be trained on various datasets to improve the effectiveness.

Therefore, several future challenges remain:

1. High-quality deepfake detection:

Most existing facial swap detection approaches are trained and evaluated on datasets that contain artifacts, such as FF++, DEDC, and DeepFake-TIMIT. While these approaches have achieved high detection accuracy, some detection systems still struggle with removing these features from manipulated facial swaps. Maintaining realistic fake imagery remains an ongoing challenge, even for high-performing detection systems. Moreover, there are datasets featuring realistic facial swaps, but no current detection system has been trained or evaluated using such these datasets. Additionally, facial swap detection systems under partial occlusions have not achieved accurate performance due to the high-quality nature of the swaps.

2. Real-time deep fake detection:

The widespread availability of deepfake generation tools has made this a crucial challenge, especially as social media platforms make it easier to generate and disseminate manipulated images and videos. Consequently, real-time deepfake detection approach has become essential for identifying manipulation techniques as soon as they are encountered. Facial swap can cause significant harm if the manipulated media is not recognized as fake before it spreads.

3. Expression swap detection:

Most existing detection systems are specifically designed to detect facial swaps using publicly available datasets. Expression swaps, however, present a unique challenge that requires improved detection methods. To address this, researchers should focus on generating comprehensive and diverse expression swap datasets. These datasets will help improve and enhance deepfake detection techniques, particularly for detecting expression swaps.

6. Conclusion

The widespread adoption of AI techniques has made facial manipulation, especially facial swaps, accessible to anyone capable of easy for creating realistic media. As the quality of deepfakes generated based on Al poses new challenges. Consequently, AI techniques have also been employed to detect such manipulations, including facial swaps. In this research, we present a comprehensive study that examines and evaluates existing generative and detection approaches based on DL. Our focus is on specific manipulations, such as facial swaps.

Additionally, we explore existing deepfake datasets and discuss the limitations and weaknesses of current DL approaches. furthermore, we highlight potential gaps in future research and provide suggestions to address these challenges. Our intention is to support the development or refinement of DL-based detection techniques specifically designed for facial-swap-based deepfakes, contributing to the preservation of traditional image media integrity and reliability.

This paper offers a new perspective on the current state of deepfake research, providing valuable insights into opportunities and challenges of deepfake detection. Our aim is to actively engage a large community of researchers in this field actively. We have nothing more, using DL methodologies, to provide more holistic understanding of deepfake landscape.

References

1. A. Siepen, "*Constructing'Liveness' on Social Media to establish'Authenticity'-BeReal, a case study* (Master's thesis)," 2023.
2. S. H. Al-Khazraji, H. H. Saleh, A. I. Khalid, and I. A. Mishkhal, "Impact of deepfake technology on social media: detection, misinformation and societal implications," *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, vol. 23, pp. 429–441, 2023.
3. I. Masi, A. Killekar, R. M. Mascarenhas, S. P. Gurudatt, and W. AbdAlmageed, "Two-branch recurrent network for isolating deepfakes in videos," 2020. https://doi.org/10.48550/arXiv.2008.03412.
4. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and beyond: A survey of face manipulation and fake detection," 2020.
5. Dixit, Priyanka, and Sanjay Silakari. "Deep learning algorithms for cybersecurity applications: A technological and status review," *Computer Science Review*, 39, 100317, 2021.
6. D. Pan, L. Sun, R. Wang, X. Zhang, and R. O. Sinnott, "Deepfake detection through deep learning," *IEEEACM Int. Conf. Big Data Comput. Appl. Technol. BDCAT*, pp. 134–143, 2020. https://doi.org/10.1109/BDCAT50828.2020.00001.
7. Mohamed R. Shoaib *et al.*, "Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models." *2023 International Conference on Computer and Applications (ICCA)*, IEEE, 2023.
8. K. Liu, I. Perov, D. Gao, N. Chervoniy, W. Zhou, and W. Zhang, "Deepfacelab: Integrated, flexible, and extensible face-swapping framework," *Pattern Recognit*, vol. 141, p. 109628, 2023. https://doi.org/10.1016/j.patcog.2023.109628.
9. I. Perov, D. Gao, N. Chervoniy, K. Liu, S. Marangonda, C. Umé, M. Dpfks, C. S. Facenheim, L. RP, J. Jiang, S. Zhang, P. Wu, B. Zhou, and W. Zhang, "DeepFaceLab: Integrated, flexible, and extensible face-swapping framework," 2021. https://doi.org/10.48550/arXiv.2005.05535.
10. Zao Download ✓ Android, iPhone, iPad 2020 [WWW Document], n.d. URL https://zaodownload.com/(accessed12.8.23).
11. FakeApp Tutorial: how to create fake videos, "2020. Malavida," URL https://www.malavida.com/en/faq/fakeapp-tutorial-how-to-create-fake-videos.
12. Olivera-La Rosa, A., Arango-Tobón, O. E., and Ingram, G. P., Swiping right: face perception in the age of Tinder, *Heliyon*, 5, no. 12, 2019.
13. M. Zendran and A. Rusiecki, "Swapping face images with generative neural networks for deepfake technology – experimental study," *Procedia Comput. Sci.*, vol. 192, pp. 834–843, 2021. https://doi.org/10.1016/j.procs.2021.08.086
14. L. Li, J. Bao, H. Yang, D. Chen, and F. Wen, "FaceShifter: *Towards high fidelity and occlusion aware face swapping*," 2020.
15. Y. Li and S. Lyu, "*Exposing deepfake videos by detecting face warping artifacts*," 2019.
16. Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm, Eds., Computer Vision – ECCV 2020, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 86–103, 2020. https://doi.org/10.1007/978-3-030-58610-2_6.
17. H. Zhao, W. Zhou, D. Chen, T. Wei, W. Zhang, and N. Yu, "Multi-attentional deepfake detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2185–2194, 2021.
18. C. Kong, B. Chen, H. Li, S. Wang, A. Rocha, and S. Kwong, "Detect and locate: Exposing face manipulation by semantic- and noise-level telltales," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 1741–1756, 2022.
19. Ju, Xingwang, "An overview of face manipulation detection," *Journal of Cybersecurity*, 2, no. 4, 197, 2020.
20. Woo, Simon, "Add: Frequency attention and multi-view based knowledge distillation to detect low-quality compressed deepfake images," In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, pp. 122–130. 2022.
21. Zi, B., Chang, M., Chen, J., Ma, X., and Jiang, Y.-G., "*WildDeepfake: A Challenging Real-World Dataset for Deepfake Detection*," 2021.
22. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., "Faceforensics++: Learning to detect manipulated facial images," In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1–11, 2019.
23. Nirkin, Yuval, Iacopo Masi, Anh Tran Tuan, Tal Hassner, and Gerard Medioni, "On face segmentation, face swapping, and face perception." In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 98–105. IEEE, 2018.
24. Thanh Thi Nguyen, Q. V. H. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, T. Huynh-The, S. Nahavandi, Thanh Tam Nguyen, Q.-V. Pham, and C. M. Nguyen, "Deep learning for deepfakes creation and detection: A survey," *Comput. Vis. Image Underst.*, vol. 223, p. 103525, 2022. https://doi.org/10.1016/j.cviu.2022.103525.
25. B. M. Smith and L. Zhang, "Joint face alignment with non-parametric shape models," in A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid, Eds., Computer Vision–ECCV 2012, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 43–56, 2012. https://doi.org/10.1007/978-3-642-33712-3_4.
26. Masood, Momina, Mariam Nawaz, Khalid Mahmood Malik, Ali Javed, Aun Irtaza, and Hafiz Malik, "Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward," *Applied intelligence*, 53, no. 4, 3974–4026, 2023.

27. *Faceswap: Deepfakes software for all*, Available at: https://github.com/deepfakes/faceswap. Accessed: September 08, 2020.
28. GitHub - dfaker/df: *Larger resolution face masked, weirdly warped, deepfake*, [WWW Document], n.d. URL https://github.com/dfaker/df (accessed 3.11.24).
29. StromWine, "*StromWine/DeepFake_tf*," 2024 .
30. GitHub - shaoanlu/faceswap-GAN, "*A denoising autoencoder + adversarial losses and attention mechanisms for face swapping*," [WWW Document], 2020. URL https://github.com/shaoanlu/faceswap-GAN (accessed 3.11.24).
31. Y. Nirkin, Y. Keller, and T. Hassner, "FSGAN: Subject agnostic face swapping and reenactment," 2019. https://doi.org/10.48550/arXiv.1908.05932.
32. Z. Xu, X. Yu, Z. Hong, Z. Zhu, J. Han, J. Liu, E. Ding, and X. Bai, "*FaceController: Controllable attribute editing for face in the wild*," 2021.
33. Y. Wang, X. Chen, J. Zhu, W. Chu, Y. Tai, C. Wang, J. Li, Y. Wu, F. Huang, and R. Ji, "*HifiFace: 3D Shape and semantic prior guided high fidelity face swapping*," 2021.
34. R. Chen, X. Chen, B. Ni, and Y. Ge, "SimSwap: An efficient framework for high fidelity face swapping," In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 2003–2011, 2020. https://doi.org/10.1145/3394171.3413630.
35. F. Rosberg, E. E. Aksoy, F. Alonso-Fernandez, and C. Englund, "FaceDancer: Pose- and occlusion-aware high fidelity face swapping," Presented at the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). *IEEE Computer Society*, pp. 3443–3452, 2023. https://doi.org/10.1109/WACV56688.2023.00345.
36. S. Agarwal, H. Farid, T. El-Gaaly, and S.-N. Lim, "Detecting deep-fake videos from appearance and behavior," In *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*, IEEE, pp. 1–6, 2020.
37. Sabir, Ekraam, Jiaxin Cheng, Ayush Jaiswal, Wael AbdAlmageed, Iacopo Masi, and Prem Natarajan. "Recurrent convolutional strategies for face manipulation detection in videos," *Interfaces (GUI)*, 3, no. 1, 80–87, 2019.
38. Daniel Mas Montserrat *et al.*, "Deepfakes detection with automatic face weighting," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020.
39. B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge (DFDC) preview dataset," 2019. https://doi.org/10.48550/arXiv.1910.08854.
40. S. Kingra, N. Aggarwal, and N. Kaur, "LBPNet: Exploiting texture descriptor for deepfake detection," *Forensic Sci. Int. Digit. Investig*, 42–43, 301452, 2022. https://doi.org/10.1016/j.fsidi.2022.301452.
41. S. Waseem, S. A. R. S. Abu-Bakar, Z. Omar, B. A. Ahmed, S. Baloch, and A. Hafeezallah, "Multi-attention-based approach for deepfake face and expression swap detection and localization," *EURASIP J. Image Video Process*, vol. 14, 2023. https://doi.org/10.1186/s13640-023-00614-z.
42. N. U. Huda, A. Javed, K. Maswadi, A. Alhazmi, and R. Ashraf, "Fake-checker: A fusion of texture features and deep learning for deepfakes detection," *Multimedia Tools and Applications*, vol. 83, no. 16, pp. 49013–49037, 2024.
43. 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*, pp. 3207–3216, 2020.
44. J. Yang, A. Li, S. Xiao, W. Lu, and X. Gao, "Mtd-net: learning to detect deepfakes images by multi-scale texture difference," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 4234–4245, 2021.
45. V. L. Thing, "Deepfake detection with deep learning: Convolutional neural networks versus transformers," In *2023 IEEE International Conference on Cyber Security and Resilience* (CSR), IEEE, pp. 246-253, 2023.
46. M. Solaiman and M. S. Rana, "Enhancing global security: A robust CNN model for deepfake video detection," In *2024 7th International Conference on Information and Computer Technologies (ICICT)*, IEEE, pp. 238–243, 2024.
47. L. Y. Gong, X. J. Li, and P. H. J. Chong, "Swin-Fake: A consistency learning transformer-based deepfake video detector," *Electronics*, vol. 13, no. 15, p. 3045, 2024.
48. J. Huang, X. Wang, B. Du, P. Du, and C. Xu, "DeepFake MNIST+: A deepfake facial animation dataset," 2021. https://doi.org/10.48550/arXiv.2108.07949.
49. A. Nagrani, J. S. Chung, W. Xie, and A. Zisserman, "Voxceleb: Large-scale speaker verification in the wild," *Computer Speech & Language*, vol. 60, p. 101027, 2020.
50. P. Korshunov Jain and S. Marcel, "Improving generalization of deepfake detection by training for attribution," *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP), Tampere, Finland*, 2021, pp. 1–6. doi: 10.1109/MMSP53017.2021.9733468.
51. Kohli, Aditi, and Abhinav Gupta. "Detecting deepfake, faceswap and face2face facial forgeries using frequency cnn," *Multimedia Tools and Applications*, 80, no. 12, 18461–18478, 2021.
52. H. Liu, X. Li, W. Zhou, Y. Chen, Y. He, H. Xue, W. Zhang, and N. Yu, "*Spatial-Phase shallow learning: rethinking face forgery detection in frequency domain*," 2021.
53. J. Wu, Y. Zhu, X. Jiang, Y. Liu, and J. Lin, "Local attention and long-distance interaction of rPPG for deepfake detection," *Vis. Comput.*, pp. 1–12, 2023. https://doi.org/10.1007/s00371-023-02833-x.
54. J. Cao, C. Ma, T. Yao, S. Chen, S. Ding, and X. Yang, "End-to-end reconstruction-classification learning for face forgery detection," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.*, pp. 4113–4122, 2022.
55. Z. Yan, Y. Zhang, Y. Fan, and B. Wu, "Ucf: Uncovering common features for generalizable deepfake detection," In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22412–22423, 2023.
56. B. Huang, Z. Wang, J. Yang, J. Ai, Q. Zou, Q. Wang, and D. Ye, "Implicit identity driven deepfake face swapping detection," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4490–4499, 2023.
57. Y. Lai, Z. Yu, J. Yang, B. Li, X. Kang, and L. Shen, "*Gm-df: Generalized multi-scenario deepfake detection*," arXiv preprint arXiv:2406.20078, 2024.
58. Patrick Kwon *et al.*, "Kodf: A large-scale korean deepfake detection dataset." *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
59. 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*, pp. 5039–5049, 2021.
60. S. K. Datta, S. Jia, and S. Lyu, "*Exposing Lip-syncing Deepfakes from Mouth Inconsistencies*," arXiv preprint arXiv:2401.10113, 2024.
61. T.-N. Le, H. H. Nguyen, J. Yamagishi, and I. Echizen, "OpenForensics: Large-Scale challenging dataset for multi-face forgery detection and segmentation In-The-Wild," *International Conference on Image Processing* (ICIP), IEEE, Anchorage,

AK, USA, pp. 3587–3591, 2021. https://doi.org/10.1109/ICIP42928.2021.9506272
62. D. Cozzolino and L. Verdoliva, "Noiseprint: A CNN-based camera model fingerprint," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 144–159, 2019.
63. M. J. Kwon, S. H. Nam, I. J. Yu, H. K. Lee, and C. Kim, "Learning jpeg compression artifacts for image manipulation detection and localization," *International Journal of Computer Vision*, vol. 130, no. 8, pp. 1875–1895, 2022.
64. H. Mareen, L. De Neve, P. Lambert, and G. Van Wallendael, "Harmonizing image forgery detection & localization: Fusion of complementary approaches," *Journal of Imaging*, vol. 10, no. 1, p. 4, 2023.
65. H. Mareen, D. Vanden Bussche, F. Guillaro, D. Cozzolino, G. Van Wallendael, P. Lambert, and L. Verdoliva, "Comprint: Image forgery detection and localization using compression fingerprints," In *International Conference on Pattern Recognition*, Cham: Springer Nature Switzerland, pp. 281–299, 2022.
66. G. Li, X. Zhao, Y. Cao, P. Pei, J. Li, and Z. Zhang, "FMFCC-V: An asian large-scale challenging dataset for deepfake detection," in: *Proceedings of the 2022 ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec '22*, Association for Computing Machinery, New York, NY, USA, pp. 7–18, 2022. https://doi.org/10.1145/3531536.3532946.
67. K. Narayan, H. Agarwal, K. Thakral, S. Mittal, M. Vatsa, and R. Singh, "DF-Platter: Multi-Face heterogeneous deepfake dataset," In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Vancouver, BC, Canada, pp. 9739–9748, 2023. https://doi.org/10.1109/CVPR52729.2023.00939.
68. Z. Yan, T. Yao, S. Chen, Y. Zhao, X. Fu, J. Zhu, ... and L. Yuan, "*Df40: Toward next-generation deepfake detection*," arXiv preprint arXiv:2406.13495, 2024.
69. Z. Yan, Y. Zhao, S. Chen, X. Fu, T. Yao, S. Ding, and L. Yuan, "*Generalizing deepfake video detection with plug-and-play: Video-level blending and spatiotemporal adapter tuning*," arXiv preprint arXiv:2408.17065, 2024.
70. C. T. Tsai, C. Y. Ko, I. Chung, Y. C. F. Wang, and P. Y. Chen, "*Understanding and improving training-free ai-generated image detections with vision foundation models*," 2024. arXiv preprint arXiv:2411.19117.
71. Y. Chen, Z. Yan, S. Lyu, and B. Wu, "${X}^ 2$-DFD: A framework for e ${X} $ plainable and e ${X} $ tendable Deepfake Detection," arXiv preprint arXiv:2410.06126, 2024.
72. Abdullah, Nibras, Israa Mishkhal, Hassan Salah, Aman Jantan, and Fadratul Hafinaz Hassan, "*Facial Forensics Detection Based on Deep Learning Approaches: Comprehensive Literature Review*," Available at SSRN 4930807.
73. Israa Mishkhal, Nibras Abdullah, and Fadratul Hafinaz Hassan Aman Jantan, "*A Review on Deepfake generation and Detection based on Deep learning: Approaches, and Future Challenges*," (2024).
74. Alhayali, Royida A., Ibrahem, Mostafa Khaled Abd Alrahman Aladamey, Mohammed Rashid Subhi, Mostafa Abdulghafoor Mohammed, Amiza Amir, and Zahraa A. Abdalkareem, "Improved artificial neural networks based whale optimization algorithm," *Iraqi Journal For Computer Science and Mathematics*, 4, no. 3, 195–202, 2023.
75. Y. M. Mohialden, S. Salman, and N. M. Hussien, "Face detection performance using CNNs and bug bonuty program (BBP)," *Iraqi Journal For Computer Science and Mathematics*, vol. 5, no. 2, pp. 59–67, 2024.
76. A. D. Jasim, "A survey of intrusion detection using deep learning in internet of things," *Iraqi Journal For Computer Science and Mathematics*, vol. 3, no. 1, pp. 83–93, 2022.
77. H. Nguyen and R. Derakhshani, "*Eyebrow Recognition for Identifying Deepfake Videos*," 2020.
78. Jung, Tackhyun, Sangwon Kim, and Keecheon Kim, "Deepvision: Deepfakes detection using human eye blinking pattern," *IEEE Access*, 8, 83144–83154, 2020.
79. Y. Nirkin, L. Wolf, Y. Keller, and T. Hassner, "DeepFake detection based on the discrepancy between the face and its context." 2020.
80. Kim, Eunji, and Sungzoon Cho, "Exposing fake faces through deep neural networks combining content and trace feature extractors," *IEEE Access*, 9, 123493–123503, 2021.
81. G. Li, X. Zhao, and Y. Cao, "Forensic symmetry for deepfakes," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 1095–1110, 2023.
82. G. Wang, Q. Jiang, X. Jin, and X. Cui, "FFR_FD: Effective and fast detection of deepfakes via feature point defects," *Inf. Sci.*, vol. 596, pp. 472–488, 2022. https://doi.org/10.1016/j.ins.2022.03.026.
83. T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning self-consistency for deepfake detection," In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV). Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Montreal, QC, Canada, pp. 15003–15013, 2021. https://doi.org/10.1109/ICCV48922.2021.01475.
84. S. Ganguly, S. Mohiuddin, S. Malakar, E. Cuevas, and R. Sarkar, "Visual attention-based deepfake video forgery detection," *Pattern Anal. Appl.*, vol. 25, pp. 981–992, 2022. https://doi.org/10.1007/s10044-022-01083-2.
85. Yu, Miaomiao, Sigang Ju, Jun Zhang, Shuohao Li, Jun Lei, and Xiaofei Li. "Patch-DFD: Patch-based end-to-end DeepFake discriminator," *Neurocomputing*, 501, 583–595, 2022.
86. H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "*Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos*," 2019.
87. H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, "On the detection of digital face manipulation," in 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Seattle, WA, USA, pp. 5780–5789, 2020. https://doi.org/10.1109/CVPR42600.2020.00582.
88. L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo, "*Face X-ray for More General Face Forgery Detection*," 2020.
89. J. Wang, Z. Wu, W. Ouyang, X. Han, J. Chen, S.-N. Lim, and Y.-G. Jiang, "*M2TR: Multi-modal Multi-scale Transformers for Deepfake Detection*," 2022.
90. Wenliang Zhao *et al.*, "Diffswap: High-fidelity and controllable face swapping via 3d-aware masked diffusion." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.