

A Review on Deepfake generation and Detection based on Deep learning: Approaches, and Future Challenges

Israa Mishkhal^{1,2}, Nibras Abdullah¹, Aman Jantan¹, Fadratul Hafinaz Hassan¹

¹ School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia

² Department of Computer Science, Sciences College, University of Diyala, Iraq.

Article Information

Article history:

Received: 22, 07, 2024

Revised: 14, 09, 2024

Accepted: 23, 09, 2024

Published: 30, 09, 2024

Keywords:

Artificial intelligence

Deep learning

Deepfake generation
Deepfake detection

Face manipulation techniques

Abstract

In recent years, applications of deepfake, particularly to achieve political, economic, or social reputation aims, have been become widespread. These applications do not require high-level professional technical skills. Also, deep learning techniques like Generative Adversarial networks (GANs) have enhanced deepfake, making it more realistic. So, several researchers are looking for developing an effective method to detect a fake image or video. This paper provides a comprehensive overview of several proposed deepfake generation approaches and the approaches used to detect any manipulation. Based on feature extraction methods, this study provides an extensive review of face manipulation, especially focusing on facial swap, re-enactment, and attribute manipulation. Additionally, the study describes all existing deepfake methods and evaluates the presented detection models based on the most effective deep learning algorithms by comparing their respective evaluation metrics. Moreover, it presents the challenges and gaps in trying to enhance and develop deepfake detection techniques. It assists readers in understanding the generation and detection of deepfake mechanisms and presents the field limitations and future works.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Nibras Abdullah

Department of Computer science, School of Computer Sciences, Universiti Sains

Malaysia, Pulau Pinang

11800, Malaysia

Email: nibras@usm.my



1. INTRODUCTION

The widespread use of digital smart devices and social media applications has resulted in the exponential growth of videos and images online. With these applications, most people can easily share their audio in cyberspace. Several deepfake applications have simultaneously enhanced deep learning approaches, enabling them to easily modify any image or video online. It has become challenges to trust online news; malicious actors are spreading fake information to target individuals and damage their reputations. Furthermore, in the post-truth era, they play significant roles in manipulating public opinion by changing the truth of some information. Videos or images must be authentic and honest when used as evidence in all sector of litigation and criminal cases. To flag the manipulated images or videos on networks, sharing files from social media requires authentication and integrity, which can be a challenging task, especially deepfakes generation. There are useful, sophisticated, and easily used tools such as Zao [1], FaceSwap [2], [3], DeepFaceLab [4]. They can manipulate a video or image with authentication and integrity. DeepFake generation can be categorized into types namely: face swap [5], Face reenactment [6], Talking face generation [7], and attribute manipulations [8], [9]. The development of deepfake technologies has started from a single GAN approach [10] [11] to high quality generation models [12], [13]. Additionally, modeling frequently incorporates Nerf [14] to improve Multiview consistency capabilities [15] [16]. Due to unethical deepfake use, it becomes necessary to develop effective deepfake detection approaches to avoid the misuse of such techniques [17].

Detection systems rely on a binary classification to distinguish between authentic or manipulated images or videos. Detection of deepfake issues requires a large dataset to train a presented detection method. Many datasets are available to aid researchers in training and testing their approaches. The public dataset VidTIMIT contains both low-and high-quality deepfake videos [18]. This dataset can effectively mimic some expressions, such as eye blinking and mouth movements. Kroshunov and Marcel produced deepfake datasets based on the GAN approach using the open-source Faceswap-GAN [19]. These datasets contain about 620 manipulated videos based GAN methods. The current deepfake research studies can be separated into two categories: deepfake generation and deepfake detection. The generation is focused on creating deep learning approaches with the least possible datasets, spending training time, and computational power. The detection presents all existing approaches that emphasize the development of robust and generic detection systems. This paper is a literature review of deepfake generation and detection from 2020 to date. It will be useful for readers to study deepfake field further in different aspects of developing deepfake generation and detection. Our contributions of this study are:

1. Presents an overview of the various types of deepfake generation approaches based on deep learning.
2. Deliver the recent deepfake detection methods based on deep learning.
3. Shows recently research gaps and opportunities in this field.

We organize the remainder of this paper as follows: Section 2 provides an overview of using deep learning algorithms for deepfake generation and summarizes the available tools. Section 3 shows all the existing deep learning approaches to detect the deepfake from 2018 to date. Then section 4 discusses the

2. DEEFAKE GENERATION APPROACHES

The technique of deepfake generation plays a significant role in traditional forgery generation methods due to removing artifacts or any manipulation traces that have been widely exploited for a detection system [20] [21]. Based on deep learning (DL), the deepfake generation has enhanced the way to extract input attributes and reconstruct them to build new manipulated images or videos with more realistic content. There are approaches to deepfake generation systems based on DL, namely: autoencoder, autoregressive [22], and GAN [23]. The type of artificial intelligent algorithms utilizes for unsupervised data representation learning. The technical aspect of its work involves converting input data into a hidden latent representation as encoder, and then reconstructing the output data, acting as a decoder. Autoencoder plays a significant role to generate deepfake tools. Figure 1 illustrates the basic of workflow of autoencoder, which train the network to extract input features and ignoring unrelated noise. This modification enables the creation of new manipulated images by generating a latent representation from the Gaussian distribution and using it as an input data in the encoder networks. The encoder operates by comparing the pixels in input data with the output data from the latent representation. Then reconstruction the output data by decoder networks [24].

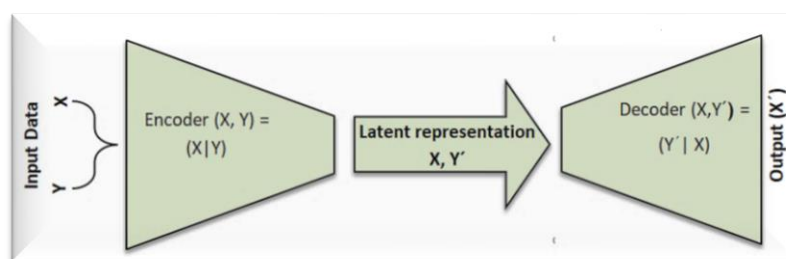


Figure 1. Autoencoder Flowchart

The autoregressive model is a statistical model, focuses on natural image distribution. The conditional distribution of each pixel depends on the previous pixels [25]. The evaluation process takes a long time to implement due to predictions and sequential evaluation processes pixel by pixel such as [26] [27]. They utilize the correlation of pixels to distinguish between manipulated and authentic images. A low correlation between pixels indicates likely manipulation of the image. The deep learning methodology known as Generative Adversarial Network (GAN) generates and enhances deepfake output. It consists of a pair of networks known as a generator network and the discriminator [28]. The aim of the generator network is to produce a new synthetic output based on distributing of input data to fool the discriminator network. On the other hand, the discriminator network aims to distinguish whether the output is real or manipulated. It utilizes to optimize the backpropagation until it reaches equilibrium between fake and real. Many software deepfake generation applications have been improved using GAN, such as FaceApp [29], Faceswap [2], ZAO [1],

RCNN for enhancing a mobile resolution (Dong: Image Super-Resolution Using Deep Convolutional), and StackedGAN for a low-quality video [30].

3. DEEPAKE GENERATION OPEN SOURCE TOOLS

There are four main categories of deepfake generation, especially for facial manipulating, namely: facial swap, facial expression, facial attribute manipulation and facial synthesis. Figure 2 shows the different types of deepfake generation for face.

1. The facial swap technique aims to switch an original face with the selecting target face with keeping the original expression [28]. In 2018, the authors presented and learned the latent spaces-based face swapping for face and hair regions using GAN approach [31]. This technique consists of two variational autoencoders for encoding face and hair regions into latent representation, as well as a deep learning methodology (GAN) for synthesizing facial swaps. The weakness of this technique is that it only applies to low resolutions (128x128). They improved their previous approach by adding deep neural network (DNN) and one variational autoencoders, as well as performing face swapping synthesis with latent variables [32]. In 2019, the researchers utilized a new deep learning methodology named recurrent neural network (RNN) to enhance the facial swap technique [33]. This technique contains three main components, namely: a unet-based recurrent reenactment generator (GR), a Pix2PixHD-based segmentation generator (GS), and a Pix2PixHD based in painting generator (GC). GR creates a mask by obtaining a pose and expression from the target, and then generating a reenacted face. GS computes the segmentation mask for the target's face and hair. Then, GC reconstructs the missing areas or any occlusion types to provide a final face swap output. To maintain the temporal coherence of face-view interpolation, they used Delaunay triangulation and barycentric coordinates. However, the resolution of the output struggles from different angles in an input image. Researches from Peking university and the Microsoft company presented the FaceShifter technique for occlusion cases in 2020 [34]. This technique utilizes embedded integration network (AEInet) and heuristic error acknowledging network (HEARnet). The AEInet denormalizes local features integration at various levels, while the HEARnet leverages the heuristic error from both input and manipulated images to identify the occlusion area. This technique presents a high level of performance during facial swapping. As summarized in Table 1, list of open sources deepfake generation tools. The famous tools for facial swapping are ZAO [1], DeepFace Lab [4], DFaker [35], Deepfakes web [8], Face Swap [2], Machine Tube [36], and Reface apps[37].

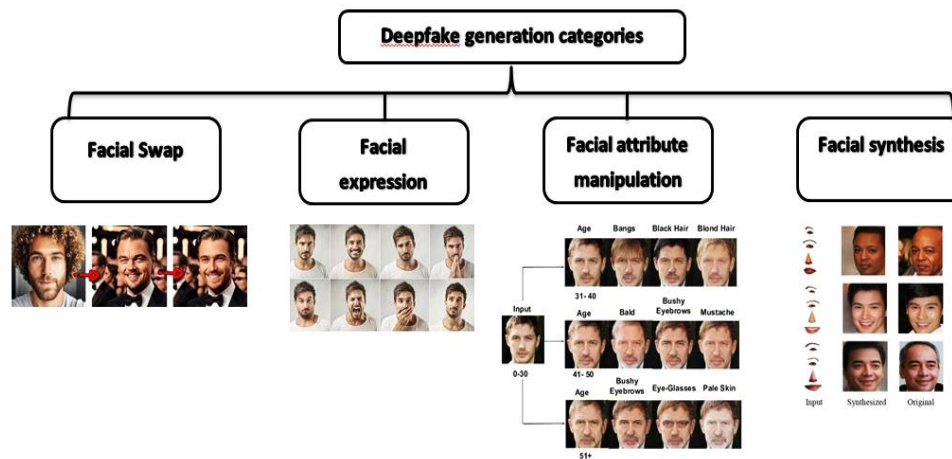


Figure 2. DeepFake generation categories

2. The facial expression is the most common deepfake generation technique because it transfers the expression from a source to the target. It is also known as the Face-to-Face technique. This technique aims to synthesize the expression of the target to show something that would never said in the real world. In 2018, Choi et al. proposed a new approach that relied on CycleGAN called StarGAN. This approach focuses on a multi-domain-based translation network using a single model. This technique makes it easier to transfer multiple expressions by supporting mask vectors for various types of facial expressions [38]. At the same year, Wu et. al. presented another approach that relied on CycleGAN to enhance mapping boundary transformation and implement an encoder-decoder (Pix2Pix) for reconstructing the synthetic output from source to target in facial expression [39]. Bansal et al. collaborated with the Facebook to propose a new approach called RecycleGAN. This method leveraged CycleGAN lost function, recycle lost, cycle consistency and adversarial loss to develop a recycle formulation that enhanced spatiotemporal contains [40]. Song et al. presented an approach to extract audio features by

applying MFCC. It utilized a conditional recurrent network that aims to temporal coherence to lip movement in an adversarial manner [41].

In 2020, Soumya et al. proposed a new approach for controlling pose and facial expression named the Interpretable and Controllable Face Reenactment network (ICface). This approach is based on two stages: the first is the facial attribute extraction of the input image, such as action unit values (AU) and interpretable head pose angles. The second stage uses GAN algorithm to integrate the extracted attributes with the input image. This approach demonstrates good performance in facial expression and poses transformation with less distortion compared to the baselines. However, it needs to enhance the noticeable artifacts in the output image [42]. Yunlian et al. introduced an approach that is based on StarGAN and CycleGAN called Ordinal Ranking Adversarial Networks (ORAN). It utilized a multi-scale discriminator and one hot label to extract the rank of the image's age and expression intensity. This approach presented the synthesis of concepts correctly based on the age and facial expression in the input image [43]. This paper [44] presented a new approach using 3D convolutional filters based on a spatial-temporal scheme to produce a high-quality deepfake video. This approach uses a static image with the desired facial expression and pose as the input for the neural network methodology. This approach struggles to transfer the facial expression and pose in high resolution due to the difficulty of controlling the probability distribution of high resolution textures. In 2021, Chaoyou et al. introduced an enhancing approach using a semi-supervised encoder-decoder based on facial expression and pose to produce the image boundary. Using LightCNN, they mapped the boundary and extraction features to encode the input image. Then, they decoded the concatenation of the boundary and the features to perform the final target synthesis. They produced a new high-resolution MVF-HQ dataset to assist future research in the same field [45]. Deepfake generation tools for transferring facial expression are Jiggy [46] and Impersonator++ [47].

3. The facial attributes involve hairstyle, eye colour, skin colour, gender, age and wrinkles. These attributes can play a significant role in altering the appearance of a person [28]. StraGAN methodology utilized for manipulating facial attributes by implementing a mask vector to support multi-domain training. In this paper [38], the authors presented this methodology as the representative domain to domain translation network. In the same year, Xiao et al. presented a new approach using CycleGAN based translation network (ELEGANT). However, the output of this approach had some artifacts [48]. To change makeup styles, Li et al. introduced BeautyGAN, which can transfer it from one person to another without facial disfigurement. They implemented pixel-level makeup loss to improve the output's realism, as well as a perceptual loss to preserve facial identity and minimize artifacts [49]. A new approach, known as AttGAN, emerged in 2019 to address the artifacts that surfaced in two previous approaches. It focuses on producing high-quality facial attribute manipulation. The classification attributes were used by the authors to ensure that facial attribute exchange was preserved during manipulation. However, it struggled to manipulate a large area of facial attributes [50]. To solve the blurry issue, Lue et al. proposed an STGAN approach that embeds a selective transfer unit with the encoder-decoder network. In the same year, this approach demonstrated better synthetic deepfake quality than the previous proposed approaches [51]. Jo et al. proposed SC_FEGAN, an approach using GAN based on free forms masks, colors or sketches. This approach produced a high-quality deepfake image with fewer artifacts because it relied on freeform feature extraction and face segmentation. They utilized a holistically nested edge detection system and histogram equalization to deal with the input sketch data [52]. In 2021, researches presented URCA-GAN network to manipulate the specific attributes of the input image differently than the target. This approach used URCAM and StarGAN. The URCAM was used to determine the attention map with the most distinctive features. In the deepfake generation, the result was high quality and fewer artifacts [53]. The papers [54] and [55] utilized lost function to keep facial attribute manipulation, such as identity, expression, and age, from different perspectives. In 2021, Affifi et al. introduced an approach to naturally change skin tone. This approach relied on StyleGAN methodology called a colour histogram based generative model (HistoGAN). It applied a color histogram for two blocks of StyleGAN [56]. The tool that used to produce facial attribute deepfake is FaceApp [27].
4. By learning the latent representation of the dataset, the facial synthesis technique generates a hyperrealistic synthetic face. The gaming technique and modelling industries utilize it to produce a virtual face. However, it becomes a dangerous technique when it used to generate a deepfake for real people to change their legal activities. In 2019, Karras et al. introduced a technique to synthesize a pose or consistent style known as ProGAN. It relies on adaptive instance normalization (AdaLN). However, this approach has artifacts in the synthesis output that are easy to detect [57]. Then in 2020, they improved their previous approach called StyleGAN. This approach successfully produced a high deepfake image quality and fewer artifacts [58]. These two approaches (ProGAN and StyleGAN) are utilized to produce a facial synthesis dataset.

Table 1: Deepfake generation & Available tools or applications

Ref.	Category	Year	Approach name	weakness	Tool/app	feature		
[31]	facial swap	2018	latent spaces-based faceswapping	Apply just on 128x128 low resolutions	ZAO DeepFace Lab DFaker Deepfakes web Face Swap Reface apps	ZAO applies face swap with celebrities from TV show or movie.		
[32]		2019	latent spaces-based DNN	Capture complex features and multimodal distributions in the latent space		DeepFace Lab applies the face swap in videos only.		
[33].			Faceswap based RNN	the resolution of the output struggles from different angles in an input image		DFaker & Deepfakes web support training of face swap model		
[34]		2020	FaceShifter	Apply for occlusion cases		Face Swap supports face swap between peoples		
[38]	facial expression	2018	StarGAN	focus on a multi-domain-based translation network using a single model	Jiggy Impersonator++			
[39]			CycleGAN					
[40]		2019	RecycleGAN					
[41]			MFCC	extract audio features only.		Jiggy allows to animate the person in static image to dance motion		
[42]			ICface	Has the noticeable artifacts in the output image		Impersonator++ support motion transfer using image synthesis		
[43]		2020	ORAN	Use just for extracting the rank of the image's age and expression intensity				
[44]			3dimation convolutional filters based on a spatial-temporal scheme	struggle to transfer the facial expression and pose in high resolution				
[45]		2021	encoder-decoder based semi-supervised					
[38]		facial attributes	2018	StraGAN			FaceApp	
[48]				ELEGANT		has some artifacts		
[49]	BeautyGAN			change only makeup styles				
[50]	2019		AttGAN	struggle to manipulate a large area	FaceApp supports modification of facial expression and face attributes			
[51]			STGAN	solve just the blurry issue				
[52]			SC_FEGAN	Has fewer artifacts in the output.				
[53]	2021		URCA-GAN	Extract only the most distinctive features.				
[56]			HistoGAN	to naturally change skin tone				
[57].	facial synthesis	2019	ProGAN	artifacts				

[58].	2020	StyleGAN	Easy to know a generated output to its source.
-------	------	----------	--

4. DEEFAKE DETECTION APPROACHES BASED ON FEATURES EXTRACTION

A deepfake detection system involves four stages to detect any manipulation in an image or video: 1) data preprocessing, 2) extracting features, 3) learning features, and then 4) classification as real or fake [28]. Learning features is a significant step in solving a complex issue in face recognition and detection. A deepfake detection system involves a block box feature extraction based on convolutional neural network (CNN) methodology. The system utilizes feature extraction to automatically learn features from the training stage. The most existing detection systems utilize specific features extracted from neural network as an input for their proposed approaches. As shown in Table 2, these are:

- a. **Biometric artifact:** the authors rely on biometric artifact as the first extraction feature. In 2018, the authors presented an approach using VGG model and LSTM to detect the time series of eye-blinking activity. However, this approach struggles to detect cases eye-blinking frequency that is unhealthy. It is difficult to deal with people who have mental illness or neurological issues [59]. Biometric eyebrow matching extraction introduced an alternative approach. It can be applied only for famous people because it relies on identity matching between the source and the target. Also, it needs big data to train [60]. A new presented approach called FakeCatcher based on CNN methodology. It emphasizes the detection system according to biological signal (PhotoPlenthysmoGraphy (PPG)). This technique can detect skin colour changing due to the peripheral circulation or blood pumping through face. This approach achieved 96% detection accuracy. However, the performance of this approach may decryes when using biased dataset for training it [61]. Mouth is another biometric artifact on which some researchers focused. In 2020, Agarwal et al. proposed an approach that aligns mouth movements with the corresponding spoken phoneme. They evaluated their approach using CNN methodology, but the manual operations required to handle phoneme and visemes alignment took time [62]. In 2021, Yang et al. presented a deepfake detection approach using lip sequences based on the person's talking habits. For pre-processing input data, they used a random password strategy. Additionally, they utilized a Dlib detector to extract the lip region. Then, they used the Connectional Temporal Classification methodology (CTC) to convert the lip region to a lip sequence. To evaluate whether the lip sequence conforms to the input data style, they used a dynamic talking habit-based speaker authentication network (SADTHnet). However, this approach depends on a lip sequence style which is difficult to find for a real person [63]. In the same year, Haliassos et al. introduced an approach based on two pre-trained networks (Resnet-18 and MS-TCN) to train pre-processed grayscale lip-cropped frames. This approach achieved a good accuracy, but it could not deal with the occlusion dataset [64].
- b. **Pixel Features:** Zhang et al. proposed an approach based on chrominance components. Using Scharr, they converted an input image from RGB to YCrCb to extract the border information. This operator turns the input image into grey-level concurrence matrix (GLCM). They utilized a convolution deep neural network to extract and classify features. However, this approach struggles when handling image resizing because it disrupts local texture and spatial correlation for deepfake detection [65]. In 2020, Khodabakhsh et al. introduced an approach for deepfake detection based on pixel features using a ResNet-CNN with a universal background model (UBM). This approach predicts the conditional probabilities for each pixel in an input image and trains on pristine data only to enhance feature extraction. This approach has limitations for uncompressed data [66]. In 2021, Zhang et al. introduced an approach that dealt with compression videos. They proposed a self-supervised decoupling network (SDNN) for learning features from the authenticity and compression input datasets. They applied this approach with varying compression rates to enhance the detection system, ensuring it remains unaffected by input compression. However, unseen compression rates affect its performance [67]. Also, Chen et al. presented an approach based on compression video input. They proposed a lightweight principal component analysis (PCA) called DefakeHop. They used PixelHop++ to extract features from varying facial regions. Additionally, they utilized subspace approximation with adjusted bias (Saab) to reduce the specific dimension of each part. However, the performance of DefakeHop effects decreases when using a lower video quality [68].

On the other hand, some papers focused on extracting frequency features to design a deepfake detection system. Li et al. proposed a system based on a frequency-aware discriminative feature learning framework (FDFL). The system effectively addressed the issue of ambiguous feature discrimination of softmax loss, as well as the low efficiency of artifact features for detection. They proposed a single-center loss (SCL) system to extract only different intra-classes of natural faces and then push the fake features. Using SCL and softmax

loss with FDFL achieved better results in the detection system, but it struggles when using unseen datasets [69]. Liu et al. relied on the properties of natural images, which consider the phase spectrum to provide extra information and complement loss from frequency components. Therefore, they proposed the Discrete Fourier Transform (DFT) to extract the phase spectrum. However, when testing this system on up-sampling dataset, its performance struggles [70].

Table 2: Summary detection approaches based on features extraction

Feature extraction	Type of feature	Ref,	Methodology	weakness
Biometric artifact	eye-blinking	[59]	VGG & LSTM	Difficult to deal with people who have mental illness or neurological issues.
	eyebrow	[60]	LightCNN, Resnet, DenseNet & SqueezeNet	Apply just for famous people & need big data to train.
	skin color	[61]	FakeCatcher-CNN	bias dataset
	mouth movements	[62]	CNN	Spend a long time
	lip sequence	[63]	Dlib, CTC & SADHnet	difficult to find dataset for a real person
	lip-cropped	[64]	Resnet-18 & MS-TCN	occlusion dataset
Pixel feature	chrominance components	[65]	CNN	disrupt local texture and spatial correlation due to image resizing
	low likelihood pixel values on the edges	[66]	RestNet-CNN & UBM	uncompressed data
	Compression video	[67]	SDNN	unseen rates
	Compression video	[68]	PCA & PixelHop++& Saab	lower video quality
	frequency features	[69]	SCL & softmax loss with FDFL	unseen datasets
	phase spectrum	[70]	DFT	up-sampling dataset
spatial-temporal	optical flow	[71]	CNN	limit to a single dataset
	heartbeat monitoring	[72]	DeepRhythm	Other factors can effect on PPG
	temporal inconsistencies	[73]	CLRNet	not fully addressed all the temporal features

- c. Spatial-Temporal Features:** Another utilized technique to extract features for detecting deepfakes is based on a spatial-temporal structure of video. Amerini et al. proposed a sequence-temporal-based approach to investigate possible dissimilarities in a video. They focused on optical flow fields to capture inter-frame correlations as input to CNN. The idea of using optical flow is to distinguish between the original video and deepfake. The evaluation of this approach is limited to a single dataset [71]. In 2020, Qi et al. introduced an approach that utilized both spatial and temporal aspects called DeepRhythm. It relied on heartbeat monitoring and dual spatial-temporal attention (Dual-ST AttenNet) to capture any dynamically changing face (PPG). They proposed a motion-magnified spatial-temporal representation (MMSTR) to produce adaptive spatial attention features. Also, to capture temporal attention, they also used the LSTM method and the Meso-4 network to extract the frame-level temporal attention, and the last step was ResNet-18 to classify these extraction features. However, PPG can be affected by other natural factors, such as sunburn and sport activity or sensitive skin [72]. In the same year, Tariq et al. proposed a new approach called the Convolutional LSTM Residual Network (CLRNet). This approach used to capture the temporal inconsistencies in a deepfake video, such as unnatural-looking artifacts or sudden changes in brightness. However, it does not fully address the temporal features of frame relations in input videos [73].

5. DEEP LEARNING DETECTION APPROACHES

Various studies have categorized deepfake detection models based on many perspectives such as the proposed methods, features extraction, and dataset type [74] [75]. This section presents the most successful approaches from 2018 to date based on deep learning. The main goal in this study is to demonstrate the most effective deep learning approaches in deepfake detection (see Table 3). Categorized deepfake detection approaches based on deep learning are:

1) Convolutional Neural Networks CNNs-based approaches are suited for detecting deepfake because they can efficiently extract various features from images and videos. Many researchers have been developing CNN-based models to enhance deepfake detection systems because CNNs can focus on identifying inconsistencies, such as unnatural eye, blinking patterns, distorted facial textures, irregular lighting, and shadowing. In 2018, Li et al. proposed a new deep learning-based method that can capture any manipulated video. Their proposed methods are based on CNNs and can effectively capture a fake video. They trained four CNN models (VGG16, ResNet50, ResNet101 and ResNet152). Some factors such as the illumination change, head motions, and face occlusions, affect the performance of this model [76]. According to CNN, Li et al. introduced another approach in the same year. Their approach consists of Long-term recurrent CNNs (LRCNs) and LSTMs to detect eye-blinking in a fake video. They considered the temporal relationship between opening the eye and closing it. Therefore, LRCN can recognize the artifacts in a single image, and LSTMs can effectively endow the ability to model long-range dependencies in sequential data. This approach relies solely on the absence of blinking as a cue for deepfake detection, but sophisticated forgers can still create realistic blinking [59]. In 2019, Amerini et al. introduced an approach that combines optical flow analysis and CNNs for detecting deepfake videos. This approach appears to be robust enough to handle video manipulations such as heading movements and changes in lighting conditions. Due to using optical flow analysis with CNNs, this approach requires a lot of processing time [71]. Another presented approach is based on taking advantage of the temporal features. The authors connected the CNN model, which has proven to be effective in detecting manipulations using the extracted temporal features from image streams. They chose two CNN architectures to build their approach, such as ResNet and DenseNet, because these improve efficiency in capturing low-level manipulation artifacts and extracting features at different levels of hierarchy. Some limitations lowered the performance of this approach, such as a limited number of training samples and unstable training of the spatial transform network (STN) [77]. Zhang et al. introduced a CNN approach based on depth wise separable convolution that improved its effectiveness in detecting deepfakes. This approach combines various levels of forensic cues in the image, such as semantic, pixel, and frequency levels. It demonstrated a good performance on several of manipulation and synthesis images. However, it tested on a limited dataset that might not fully encompass the range of deepfake techniques [65]. Zhuang et al. proposed an approach that consists of coupled network architecture with two steps pairwise learning for detecting GAN images. This approach's architecture consists of two DenseNet sub-networks. These sub-networks are trained to learn complementary features, which enhances their performance. The approach achieves a good deepfake detection accuracy. However, it tested just on images generated by specific GAN methods [78]. In 2020, Chung et al. proposed a DenseNet approach that uses a contrastive loss called CFFN. This approach evolved into a two-streamed network structure that accepts pairwise datasets as an input. It can capture the discriminative features of fake images. When the input image differs from used training data used, the CFFN approach cannot detect any manipulation [79]. Kumar et al. introduced a deep learning-based approach for detecting deepfakes, which are generated using pixel features. This approach is based on five parallel ResNet-18 networks connecting with RGB frames for extracting localization facial artifacts and noise patterns. Additionally, the approach performs poorly when detecting heavily compressed frames. This approach utilized five parallel ResNet-18 networks, thus increasing the computational complexity [80]. Another deep learning-based approach is DeepRhythm, which utilizes ResNet and LSTM methodologies. Qi et al. introduced it in 2020, relying on heartbeat monitoring and spatial-temporal attention. It achieved high performance; however, it has some factors that affect its performance, such as sports activities, sensitive skin, and sunburn effects [72]. Nguyen et al. introduced four deep-learning approaches to detecting eyebrows. These approaches are LightCNN, Resnet, DenseNet, and SqueezeNet. They conducted two experiments, one short-term and one long-term, to evaluate the similarity between reference and probe eyebrows. This approach achieved an accuracy of 87.9% using ResNet on the right eyebrow, while the worst obtained result was 69% using DenseNet on the same eyebrow. These results show that using eyebrow was not a perfectly chosen feature to detect deepfakes [60]. The authors proposed another deep-learning model that uses CNNs to detect the manipulation of images. The authors frame the deepfake issues as a binary classification problem. In their classifier approach, they utilized the attention mechanism to process the feature map. It can highlight the manipulation pixels to guide the CNN networks to detect these regions. This approach shows high performance over all existing models [81].

Wang et al presented a novel approach based on monitoring neuron behaviours to detect manipulated faces called FakeSpotter. The authors utilized mean neuron coverage (MNC) to extract behaviours of each neuron activation layer. CNN networks frequently employ MNCs to investigate their internal behaviours. The fakeSpotter technique is utilized to identify four types of manipulated facials: entire synthesis, attribute editing, expression manipulation, and deepfake. This approach achieved excellent performance in detecting various types of deepfakes. However, it has some limitations in detecting swapping faces and voices because FakeSpotter focuses only on facial images without considering voice [82]. Li et al. proposed a new patch-and-pair CNN (PPCNN) approach that focuses on learning the complete features of faces. PPCNN has a two-branch learning framework: 1) extracting different features between real and fake patches; 2) capturing the inconsistencies between the region of the face and the region around the face. The PPCNN performance improves accuracy compared to the previous approaches; however, PPCNN relied on comparing and capturing the difference between the face region and its background. Therefore, an input image containing only a face region may result in reduced performance. In 2021, researchers utilized facial details, including the combination of direct light and identity texture, to introduce a new XceptionNet based approach known as forget-detection-with-facial-detail (FD2Net). The 3D decomposition of the face image, which includes identity and common textures, ambient light, and direct light, disentangles this approach from the input image. The authors of this study found critical forgery clues in identity texture and direct light. Then they showed the manipulated region by using UV space and brought out the subtle forgery patterns. This approach achieved state-of-the-art performance; however, it is limited by relying on specific datasets [83].

A frequency-aware discriminative feature learning framework (FDL) was created by Li et al. to deal with the problems of softmax loss and low efficiency of artifacts that have unclear frequency features. They proposed a single center loss (SCL) to extract real features and remove the manipulated features. However, using unseen datasets affected the performance of this approach [69]. Luo et al. presented an Xception with an SRM approach to capture a high-frequency noise feature of manipulated datasets. This approach consists of three models to take full advantage of the high-frequency features. These models include the extraction of multi-scale high-frequency features, an attention dual cross-modality, and an attention residual guided spatial model. This model achieved state-of-the-art performance when compared to other approaches, but its performance suffered when testing on heavier compression datasets [84]. Liu et al. proposed an approach that relies on the phase spectrum known as the discrete Fourier transform (DCT). This approach is utilized to extract the phase spectrum for the detection of deepfakes, as it is sensitive to up-sampling. The authors assumed that the local textual content had an impact on the detection system. This approach improves the deepfake performance, but it might be vulnerable to not up-sampling dataset [70].

In 2022, Gowda et al. suggested three approaches based on CNN: ResNext, Xception, and an ensemble of both ResNext and Xception. The result shows that the ensemble model achieved better performance than others [85]. Raza et al. proposed a hybrid approach to detecting deepfakes based on combining VGG16 and CNN architectures. This approach relies on processing pixel data. It aims to learn patterns from a historical input image to predict unseen manipulations. They employed CNN techniques to analyse a confusion matrix from a time series. This approach has limitations in detecting occlusion images [86]. Ismail et al. proposed an approach based on two different methods to detect face manipulation. They introduced CNN-based HOG (the Histogram of Oriented Gradient Method) and XceptionNet, which extract feature sets and feed them into gated recurrent units (GRU) to capture spatial and temporal features. The experimental result shows the state-of-the-art performance over other methods. However, it needs an improvement to achieve a higher detection level and discover multimodal deepfake videos [87].

Also, Awotunde et al. introduced a CNN-based approach to extract face area from video frames. They utilized ReLU with CNN to capture the discriminant spatial features. This method relies on the identifying artifacts in video frames. Occlusion and punctual movement blur affect its performance [88]. Patel et al. developed a CNN architecture approach known as dense CNN (D-CNN). By extracting each frame of the video to detect and crop faces, the authors extended this approach to classify both images and videos. The model used a data-driven approach for deepfake detection that predicts the respective class of an input image based on feature maps. It was excellent at detecting deepfakes in various dataset at low resolution. However, the limitations of this approach are the input image size and high resolution [89]. In 2024, Heidari et al. introduced a new approach based on the blockchain technique known as BFLDL. They combined SeCaps and CNN methodologies to extract feature sets from an input deepfake image. This approach utilized two distinct strategies to extract the features defined: a texture-based analysis tool to capture the final facial attributes, and SeCaps-CNN to capture spatial relationships in images with structural information. The performance enhances accuracy; however, it has limitations on a low resolution and occlusion image [90].

2) Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTMs) based approaches:

RNNs and LSTMs are two other deep-learning techniques used to detect deepfake generation. RNN frequently employs temporal discrepancies to extract features. In 2019, Sabir et al. proposed an approach that relies on exploiting temporal features from an input dataset across domains. They tested their approach on

special cases of manipulation, such as Face2Face and FaceSwap in video frames. They discovered that preprocessing input faces can improve the training performance. Also, Training a sequence of input datasets provided better performance than a single input frame. They used the DenseNet technique in training stage and RNNs for detection at different levels of the network hierarchy. This approach improves the detection accuracy by up to 4.55%. however the limitation of this approach is the lack of modeling the temporal relationship between different facial areas [77]. In 2020, Mahra introduced an approach that combines capsule networks and LSMTs. This study utilizes the spatiotemporal hybrid model to extract input data features through capsule networks, which then feeds the temporal features across the video frames to the LSMTs. This approach achieved better performance with equal interval frame selection. However, the approach encounters limitations when selecting video frames based on modification levels, and using a single frame for selection. Minor or consistent discrepancies across the selection of video frames affect its performance [91]. Also, Amerini et al. and Masi et al. introduced two approaches based on long short-term memory (LSTM). Amerini and Caldelli 2020 presented the first model to capture the temporal correlation inter-frame prediction errors that distinguish a real video from a manipulated one [92]. The second model used the Laplacian of Gaussian (LoG) to extract the color domain and frequency to highlight the artifacts in frames, which were then fed to LSTM for training and classification based on time series [93]. Sun et al. proposed an approach in 2021 that uses two-stream RNN networks to extract geometric features from capturing faces using Dlib. All these approaches tested on the FF++ dataset and achieved great results; however, their performances dropped when tested on other datasets, demonstrating their inability to generalize models, and the complicated calibration process involved in extracting features may make them hard to duplicate [94].

3) Generative Adversarial Networks (GANs) based approach:

Generative adversarial networks (GANs) for deepfake detection systems have become a significant advancement in the field of forensics. Through their unique training stage, GANs generate realistic synthetic deepfake datasets, which they then use to identify and analyze any manipulation. GAN-based detection approaches can effectively learn to detect any manipulations in an image or video by leveraging a dual-network framework that consists of a generator and discriminator. The discriminator can detect the manipulated images produced by the generator. Also, it can identify artifacts and inconsistencies during the deepfakes creation process [57]. Nguyen et al used the adversarial training stage not only to enhance an approach's ability to detect a deepfake, but also to improve the generalization capabilities to detect unseen artifacts [95]. In 2021, Aduwala et al. presented an approach based on GAN discriminators. It involves leveraging modified GANs to detect a manipulated video. This approach achieved high accuracy by analyzing facial gestures, behaviors, and facial appearance, and it can effectively detect a fake video. Compared to other traditional detection systems that rely on pixel inconsistencies or artifacts, the performance of the GANs discriminator with deepfake detection has improved significantly. However, the GAN discriminators have performance effects on unknown dataset sources [96]. Huang et al. proposed a GAN-based approach for robust localization of face manipulations named FakeLocator. They leveraged the imperfections of upsampling in GAN with a gray-scale fakeness map to capture the fake textures. Upsampling is used to detect and localize fake regions. This method outperforms other existing state-of-the-art approaches using different datasets, including the GAN-based dataset. Also, it improves deepfake detections against real-world image issues, such as low-resolution, blur, noise, and compression. Some rasing issues with FakeLocator include its inability to recognize non-additive noise adversarial attacks, the inability to reconstruct deepfake generation methods, and its inability to detect the fake texture in each image and classify it using different GANs and up-sampling methods [97]. Preeti et al. proposed a GAN-based approach to improve deepfake detection in social media, known as Deep Convolution GAN. This method relies on noise to ensure the diversity of data distribution. Also, it tested on fewer images under controlled conditions by optimizing some factors, such as the normalized batch size of images, a sufficient number of epoch cycles, effective model layers, and noise value in manipulated images [98]. In 2024, Sharma et al. combined GANs and CNN in an ensemble approach known as GAN-CNN Ensemble to detect manipulations on social media images. This approach aims to minimize catastrophic damage and enhance robustness against various deepfake techniques [99].

Table 3: Summary detection approaches based on Deep learning approaches

Ref	Year	Features selection	Method	Dataset	perfromance	weakness
CNN based approaches						
[76]	2018	Artifact	CNN(VGG&ResNet)	UADFV DeepfakeTIMIT	83.3%,97.4% 84.6%, 99.9%	the illumination change, head motions, and face occlusions
[59].		eye-blinking	LRCNs & LSTMs	CEW EBV	99%	rely solely on the absence of blinking

[71].		optical flow	CNNs	FaceForensics++	81%	long processing time
[77].		temporal features	ResNet DenseNet	FaceForensics++	94.9 % 96.9%	a limited number of training samples and unstable training of the spatial transform network (STN)
[65]	2019	depthwise separable convolution	CNN	CASIA, GPIR, COVERAGE, BigGANs, LSUN Bedroom, PGGAN, SNGAN, StyleGAN	97.5%	Limit training dataset
[78]		pairwise learning for detecting GAN images	DenseNet	GAN-Generated Images based On CelebA	98.6%	Don't test it on other types of deepfake techniques
[79]		pairwise learning	CFFN based on DenseNet	CelebA ILSVRC12	98.8%	Unable to identify a new feature that differs from the ones used in training
[80]		pixel features	ResNet-18	No-compression Easy-compression Hard-compression	99.96% 99.10% 91.20%	Has struggle with heavily compressed datasets and the computational complexity.
[72]	2020	spatial-temporal	ResNet & LSTM	DFD DF F2F FS ALL DFDC	97.5% 99.7 % 98.9 % 97.8 % 98 % 64.1%	Natural factors can affect on its performance
[60].		eyebrow	LightCNN Resnet DenseNet SqueezeNet	Celeb-DF	69.6 % 87.9 % 69% 80.2%	Eyebrow matching is not the best evaluation.
[81]		the feature map	CNNs	DFFD	99.7%	
[82]		neuron behaviors	CNN & MNC	Celeb-DF	98.6%	detect swapping faces and voice
[100]		patch level.	ResNet18 (PPCNN)	Faceforensics Mesonet DeepfakeTIMIT	99.4% 81.5% 97.8%	the difference between the face region and its background
[83]		facial details	XceptionNet (FD2Net)	F2F FS	98.22% 86.54%	Test on specific datasets
[69]		frequency	Xception (FD2Net)	FF++	99.43%	use unseen datasets
[84]	2021	high-frequency noise	Xception & SRM	DF F2F FS CelebDF. The metric is AUC.	98.6% 95.7% 92.9% 79.4%	heavier compression datasets
[70]		phase spectrum	Xception (DCT)	FF++ Celeb-DF	96.91% 76.88%	not up-sampling dataset and entirely different type of manipulated facial
[85]			ResNext, Xception, ensemble of both	DFDC	93%	Limit only two models: ResNext & Xception.
[87]	2022	spatial and temporal features	CNN-based HOG & XceptionNet & GRU	CelebDF	95.53%	computational efficiency and generalization
[86]		confusion matrix from a time series	VGG16 & CNN	Photoshopped real and fake faces	94%	occlusion images
[88]		spatial features	ReLU with CNN	DeepFake F2F	98% 95%	occlusion & punctual movement blur
[89]	2023	data-driven	D-CNN	AttGAN GDWCT StyleGAN, StyleGAN2 StarGAN	98.33% 99.33% 95.33% 94.67% 99.17%	Limit image size and high resolution
[90]	2024	blockchain	SeCaps-CNN (BFLDL)	FF++ DeepFakeTIMIT DFDCpre CelebDF	97% 97% 98.1% 98.9%	low resolution & occlusion

RNNs & LSTMs based approach						
[77]	2019	temporal features	DenseNet & RNNs	FF++(Deepfake Face2Face FaceSwap)	96.9% 94.35% 96.3%	the temporal relationship between different facial areas
[91]		spatiotemporal	CapsuleNet + LSTM	DFDC	83.42% on equal interval	modification levels & using a single frame & Minor or consistent discrepancies across the selection
[92]	2020	temporal correlation inter-frame prediction errors	LSTM	FF++	94.3%	Artifacts in dataset
[93]		the color domain and frequency on time series	CNN&LSTM	FF++	94.3%	Artifacts in dataset
[94]	2021	geometric features	RNN	FF++	99.9%	Artifacts in dataset
GANs Based approach						
[96]	2021	pixel inconsistencies & artifacts	GAN discriminators	StyleGAN DFDC	92% 66.2%	unknown dataset sources
[97]	2022	fake textures	FakeLocator	CelebA	93.04%	non-additive noise adversarial attacks, new deepfake generation methods, and the fake texture in each image
[98].	2023	Noise features	Deep Convolution GAN	celebA	99%	The Limited dataset, a small image size, noise and limited number of epoch cycles.
[99]	2024	catastrophic damage	GAN-CNN Ensemble	social media images	98.67%	Train the model on different datasets, and the inability to alter the distribution in real-word images.

6. DISCUSSION

In recent years, deepfake technology has made significant advancements and challenges. The deepfake generation technique utilizes powerful deep learning methodologies, including GANs, autoencoders, and autoregressive, to produce images or videos that are remarkably realistic. This technology has potential applications in the education, entertainment, and creative industries. However, it poses substantial ethical, societal, and security risks.

During **the deepfake generation stage**, experts create numerous high-quality images or videos in various manipulation categories, such as facial swaps, facial expressions, facial attributes, and facial synthesis. The fake generation has made significant strides. Some of these methods include self-supervised training, the use of pix2pixHD, AdanIN, feature disentanglement, and self-attention models to enhance facial swapping, the application of temporal discriminators and optical flow estimation in the manipulated videos, the reduction of artifacts through the addition a secondary network for seamless composites blending, the use of loss functions to capture occlusion, conversion, pose, illumination, and the integration of perceptual loss with the VGG network, among others. These changes have made it harder for humans to distinguish between manipulated and real content. Moreover, some limitations exist in the deepfake generation, such as the use of frontal poses in facial expressions that restrict performance, and the requirement for a perfect face match required in facial swap techniques. Based on the discussed studies, Figure 3 shows the implementation frequency of the major deepfake generation categories over the years. Most existing deepfake generations demonstrate that changing facial expressions and facial attributes are among the pioneering categories. This development has pros and cons for society due to anyone can easily acquire this technology through available software application. Moreover, most studies utilize a stable GAN structure network for deepfake generations, as it produces sharper deepfakes compared to autoencoder or autoregressive methods. However, GANs require a large dataset and a long time for training. Since deepfakes can easily ruin a person or an organization by dominating falsified data, it is crucial to learn about the various types of deepfakes and to develop a new detection approach that can help distinguishing between fake and real content.

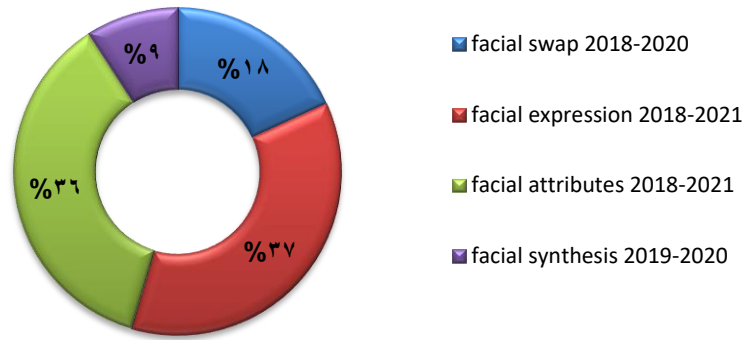


Figure 3. the major deepfake generation categories from year 2018 to 2021

In the deepfake detection stage, researchers have increasingly utilized deep learning techniques for manipulation, as the quality of deepfake generation that yields realistic images or videos is challenging for handcrafted feature extraction. In previous analyzed studies, many deep-learning techniques were proposed as ways to detect various types of manipulation. These include Convolutional Neural networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Generative Adversarial Networks (GANs). Figure 4 illustrates the prevalent deep-learning techniques utilized in recent times for deepfake identification and feature extraction.

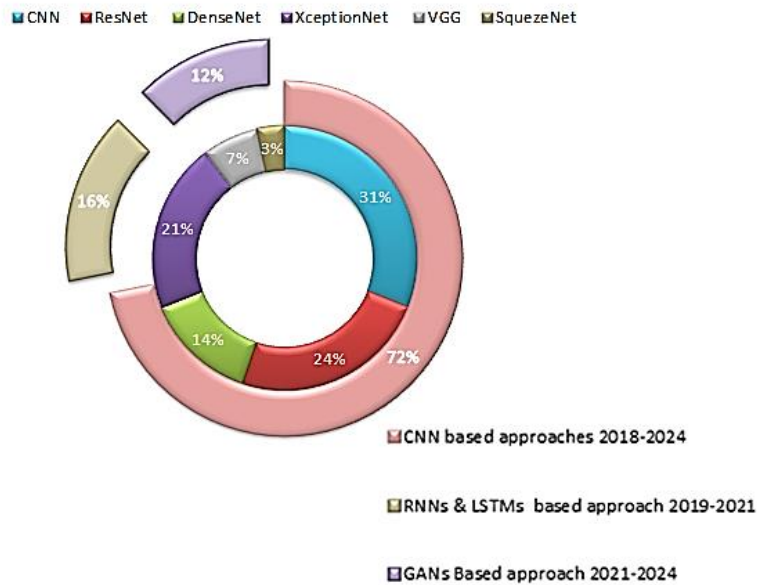


Figure 4. The implementation of deep learning for deepfake detection approaches based on discussed studies.

In recent years many studies for deepfake detection systems have utilized Generative adversarial networks (GANs) because their discriminator have improved significantly over traditional methods that relied on pixel inconsistencies or artifacts. In 2021, the GAN discriminators achieved an accuracy of 92% on the styleGAN dataset and 66% on DFDC. In 2022, the FakeLoctor system achieved about 93% accuracy on CelebA. In 2023 and 2024, some researchers introduced a hyper-approach based on CNNs and GAN. They achieved more than 97%. However, they tested on limited datasets (as shown in Table 3).

7. FUTURE DIRECTIONS

Deepfake technology is gaining a lot of attention due to its potential effects on society. This section discusses the potential opportunities and future direction in both deepfake generation and detection approaches:

1. **Deepfake generation:** most deepfakes use GANs techniques to improve the output quality, but there are still some areas that require attention, including:

- a) **Enhancing Quality in Real-Time Creation:** In real-time, deepfake generation uses image animation techniques. Consequently, the output might contain some biometric artifacts with low fidelity for certain head movements or facial expressions [101].
 - b) **Enhancing the Quality of Deepfake Generation:** Improving output quality under occlusions and varying illumination conditions is a key trend in deepfakes generation. In these scenarios, most deepfake outputs exhibit subtle traces, fingerprints, and pixel inconsistency due to abrupt change in illumination or inconsistent facial features caused by any type of occlusions [102], [103], [104].
 - c) **Enhancing the temporal coherence quality** is a significant limitation of deepfake generation due to the presence of noticeable artifacts, such as jitter and flickering between moving video frames. These issues have recently arisen because frameworks process each frame independently while neglecting temporal consistency [105], [106].
2. **Deepfake detection:** Despite the improvement in the performance of deepfake detection approaches, the development of deepfake generation remains a significant concern. Therefore, they still need to address several challenges in detecting deepfakes, including:
 - a) **Real-Time detection:** This presents a research opportunity to create a user-friendly deepfake detection method that can work in tandem with AI deepfake generation or social media platforms. Due to the massive amount of data shared every second on social media, detecting deepfakes in real time. requires the computational power, making this a difficult challenge that
 - b) **Complexity features:** The majority of existing detection systems, such as artifacts, have focused on specific types of features. As AI techniques for creating deepfakes improve, they need to develop an approach that can handle the complexity of the diverse range of deepfake features. One way to do this is to combine anomaly and signature-based ensemble learning, which will help improve the performance of deepfake detection systems.
 - c) **Temporal Coherence in Multi-Frames:** Several facial detection studies focus on individual frames in a manipulated video, neglecting the temporal consistency of sequencing the data. Utilizing temporal coherence in multi-frames can enhance deepfake detection techniques by analyzing frame sequences for inconsistencies. The majority of the challenge lies in balancing the trade-off between temporal coherence and deepfake detection accuracy. This necessitates the development of a hybrid model that can integrate both spatial and temporal features, enhancing the performance of detection systems without compromising speed or accuracy.

8. CONCLUSION

Deepfakes play a crucial role in our society, and as the deepfake generation approaches becoming more sophisticated, detecting them becomes increasingly challenging. This paper has provided an overview of deepfake generation techniques and their tools, as well as the latest advances in detection techniques based on deep learning approaches. The aim is to stay one step ahead in the race with generative intelligence, curb the spread of the fake data, safeguard the integrity of content, and address the issues that deepfakes can cause in political, economic, and personal contexts. Additionally, it presented a detailed analysis of existing deepfakes, with a particular focus on facial swap, re-enactment, and attribute manipulation generation and detection approaches, highlighting their strengths and weaknesses. It also addressed challenges and future directions for both generation and detection techniques. Therefore, there is a need for more studies to enhance detection systems that can mitigate the risks and dangers associated with the spread of fakes. Finally, deepfake techniques present opportunities and challenges that necessitate addressing common issues.

REFERENCES

- [1] "Zao Asian Cafe," App Store. Accessed: Mar. 11, 2024. [Online]. Available: <https://apps.apple.com/us/app/zao-asian-cafe/id1530895491>
- [2] shaoanlu, shaoanlu/faceswap-GAN. (Mar. 09, 2024). Jupyter Notebook. Accessed: Mar. 11, 2024. [Online]. Available: <https://github.com/shaoanlu/faceswap-GAN>
- [3] 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, Oct. 2020, pp. 2003–2011. doi: 10.1145/3394171.3413630.
- [4] K. Liu, I. Perov, D. Gao, N. Chervoniy, W. Zhou, and W. Zhang, "Deepfacelab: Integrated, flexible and extensible face-swapping framework," Pattern Recognition, vol. 141, p. 109628, Sep. 2023, doi: 10.1016/j.patcog.2023.109628.
- [5] AncilottoAlberto, PaissanFrancesco, and FarellaElisabetta, "XimSwap: Many-to-Many Face Swapping for TinyML," ACM Transactions on Embedded Computing Systems, May 2024, doi: 10.1145/3603173.
- [6] S. Bounareli, C. Tzelepis, V. Argyriou, I. Patras, and G. Tzimiropoulos, "HyperReenact: one-shot reenactment via jointly learning to refine and retarget faces," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 7149–7159. Accessed: Jun. 05, 2024. [Online]. Available: http://openaccess.thecvf.com/content/ICCV2023/html/Bounareli_HyperReenact_One-Shot_Reenactment_via_Jointly_Learning_to_Refine_and_Retarget_ICCV_2023_paper.html
- [7] A. Mir, E. Alonso, and E. Mondragón, "DiT-Head: High-Resolution Talking Head Synthesis using Diffusion Transformers," arXiv preprint arXiv:2312.06400, 2023, Accessed: Jun. 05, 2024. [Online]. Available: <https://arxiv.org/abs/2312.06400>
- [8] M. Masood, M. Nawaz, K. M. Malik, A. Javed, and A. Irtaza, "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward," Nov. 22, 2021, arXiv: arXiv:2103.00484. doi: 10.48550/arXiv.2103.00484.











- [9] Y. Pang et al., "Dpe: Disentanglement of pose and expression for general video portrait editing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 427–436. Accessed: Jun. 05, 2024. [Online]. Available: http://openaccess.thecvf.com/content/CVPR2023/html/Pang_DPE_Disentanglement_of_Pose_and_Expression_for_General_Video_Portrait_CVPR_2023_paper.html
- [10] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks".
- [11] I. Goodfellow et al., "Generative adversarial nets," Advances in neural information processing systems, vol. 27, 2014, Accessed: Jun. 05, 2024. [Online]. Available: <https://proceedings.neurips.cc/paper/5423-generative-adversarial-nets>
- [12] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Advances in neural information processing systems, vol. 33, pp. 6840–6851, 2020.
- [13] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10684–10695. Accessed: Jun. 05, 2024. [Online]. Available: http://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html
- [14] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: representing scenes as neural radiance fields for view synthesis," Commun. ACM, vol. 65, no. 1, pp. 99–106, Jan. 2022, doi: 10.1145/3503250.
- [15] K. Jiang, S.-Y. Chen, F.-L. Liu, H. Fu, and L. Gao, "NeRFFaceEditing: Disentangled Face Editing in Neural Radiance Fields," in SIGGRAPH Asia 2022 Conference Papers, Daegu Republic of Korea: ACM, Nov. 2022, pp. 1–9. doi: 10.1145/3550469.3555377.
- [16] Z. Yu, Z. Yin, D. Zhou, D. Wang, F. Wong, and B. Wang, "Talking head generation with probabilistic audio-to-visual diffusion priors," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 7645–7655. Accessed: Jun. 05, 2024. [Online]. Available: http://openaccess.thecvf.com/content/ICCV2023/html/Yu_Talking_Head_Generation_with_Probabilistic_Audio-to-Visual_Diffusion_Priors_ICCV_2023_paper.html
- [17] C. Tan, Y. Zhao, S. Wei, G. Gu, and Y. Wei, "Learning on gradients: Generalized artifacts representation for gan-generated images detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 12105–12114. Accessed: Jun. 05, 2024. [Online]. Available: http://openaccess.thecvf.com/content/CVPR2023/html/Tan_Learning_on_Gradients_Generalized_Artifacts_Representation_for_GAN-Generated_Images_Detection_CVPR_2023_paper.html
- [18] Ahmed, Hasan Maher, and Manar Younis Kashmola. "Performance Improvement of Generative Adversarial Networks to Generate Digital Color Images of Skin Diseases." Iraqi Journal of Science (2023): 4791-4805.
- [19] P. Korshunov and S. Marcel, "Vulnerability assessment and detection of Deepfake videos," in 2019 International Conference on Biometrics (ICB), Crete, Greece: IEEE, Jun. 2019, pp. 1–6. doi: 10.1109/ICB45273.2019.8987375.
- [20] N. Kanwal, A. Girdhar, L. Kaur, and J. S. Bhullar, "Detection of Digital Image Forgery using Fast Fourier Transform and Local Features," in 2019 International Conference on Automation, Computational and Technology Management (ICACTM), London, United Kingdom: IEEE, Apr. 2019, pp. 262–267. doi: 10.1109/ICACTM.2019.8776709.
- [21] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," IEEE J. Sel. Top. Signal Process., vol. 14, no. 5, pp. 910–932, Aug. 2020, doi: 10.1109/JSTSP.2020.3002101.
- [22] "A. van den Oord, Y. Li, O. Vinyals, Representation... - Google Scholar." Accessed: Jun. 07, 2024. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=A.+van+den+Oord%2C+Y.+Li%2C+O.+Vinyals%2C+Representation+Learning+with+Contrastive+Predictive+Coding%2C+arXiv+e-prints+%282018%29+arXiv%3A1807.03748.&btnG=
- [23] Goodfellow et al., "Generative adversarial networks," Communications of the ACM, Oct. 2020, doi: 10.1145/3422622.
- [24] D. P. Kingma and M. Welling, "An Introduction to Variational Autoencoders," MAL, vol. 12, no. 4, pp. 307–392, Nov. 2019, doi: 10.1561/22000000056.
- [25] "Jianchang Mao and Anil K Jain. Texture classification... - Google Scholar." Accessed: Jun. 07, 2024. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Jianchang+Mao+and+Anil+K+Jain.+Texture+classification+and+segmentation+using+multiresolution+simultaneous+autoregressive+models.+Pattern+recognition%2C+25%282%29%3A173%E2%80%9393188%2C+1992.&btnG=
- [26] Azawi, Raghad Majeed, Ibrahim Tariq Ibrahim, and Israa Mishkhal. "A Hybrid Detection System of Heart Disease by Using Machine Learning Techniques." Journal homepage: <https://ijas.uodiyala.edu.iq/index.php/IJAS/index> ISSN 3006: 5828.
- [27] "Adversarial-learning-based image-to-image transformation: A survey - ScienceDirect." Accessed: Jun. 07, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231220310559>
- [28] J. W. Seow, M. K. Lim, R. C. W. Phan, and J. K. Liu, "A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities," Neurocomputing, vol. 513, pp. 351–371, Nov. 2022, doi: 10.1016/j.neucom.2022.09.135.
- [29] "FaceApp Inc, Faceapp (2016). <https://www.faceapp.com/>. - Google Scholar." Accessed: Jun. 08, 2024. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=FaceApp+Inc%2C+Faceapp+%282016%29.+https%3A%2F%2Fwww.faceapp.com%2F.&btnG=
- [30] J. He, J. Zheng, Y. Shen, Y. Guo, and H. Zhou, "Facial Image Synthesis and Super-Resolution With Stacked Generative Adversarial Network," Neurocomputing, vol. 402, pp. 359–365, Aug. 2020, doi: 10.1016/j.neucom.2020.03.107.
- [31] R. Natsume, T. Yatagawa, and S. Morishima, "RSGAN: Face Swapping and Editing using Face and Hair Representation in Latent Spaces," in ACM SIGGRAPH 2018 Posters, Aug. 2018, pp. 1–2. doi: 10.1145/3230744.3230818.
- [32] R. Natsume, T. Yatagawa, and S. Morishima, "FSNet: An Identity-Aware Generative Model for Image-based Face Swapping," vol. 11366, 2019, pp. 117–132. doi: 10.1007/978-3-030-20876-9_8.
- [33] Y. Nirkin, Y. Keller, and T. Hassner, "FSGAN: Subject Agnostic Face Swapping and Reenactment," Aug. 16, 2019, arXiv: arXiv:1908.05932. doi: 10.48550/arXiv.1908.05932.
- [34] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen, "FaceShifter: Towards High Fidelity And Occlusion Aware Face Swapping," Sep. 15, 2020, arXiv: arXiv:1912.13457. Accessed: Dec. 10, 2023. [Online]. Available: <http://arxiv.org/abs/1912.13457>
- [35] "GitHub - dfaker/df: Larger resolution face masked, weirdly warped, deepfake." Accessed: Mar. 11, 2024. [Online]. Available: <https://github.com/dfaker/df>
- [36] "machine.tube." Accessed: Jun. 09, 2024. [Online]. Available: <http://ww1.machine.tube/>

- [37] A. Roohi, S. Angizi, P. Navaeilavasani, and M. Taheri, "ReFACE: Efficient Design Methodology for Acceleration of Digital Filter Implementations," in 2022 23rd International Symposium on Quality Electronic Design (ISQED), Apr. 2022, pp. 1–6. doi: 10.1109/ISQED54688.2022.9806144.
- [38] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 8789–8797. doi: 10.1109/CVPR.2018.00916.
- [39] W. Wu, Y. Zhang, C. Li, C. Qian, and C. C. Loy, "ReenactGAN: Learning to Reenact Faces via Boundary Transfer," presented at the Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 603–619. Accessed: Jun. 10, 2024. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Wayne_Wu_Learning_to_Reenact_ECCV_2018_paper.html
- [40] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, "Recycle-GAN: Unsupervised Video Retargeting," presented at the Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 119–135. Accessed: Jun. 10, 2024. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Aayush_Bansal_Recycle-GAN_Unsupervised_Video_ECCV_2018_paper.html
- [41] Y. Song, J. Zhu, D. Li, X. Wang, and H. Qi, "Talking Face Generation by Conditional Recurrent Adversarial Network," Jul. 25, 2019, arXiv: arXiv:1804.04786. doi: 10.48550/arXiv.1804.04786.
- [42] S. Tripathy, J. Kannala, and E. Rahtu, "ICface: Interpretable and Controllable Face Reenactment Using GANs," in 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA: IEEE, Mar. 2020, pp. 3374–3383. doi: 10.1109/WACV45572.2020.9093474.
- [43] Y. Sun, J. Tang, Z. Sun, and M. Tistarelli, "Facial Age and Expression Synthesis Using Ordinal Ranking Adversarial Networks," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2960–2972, 2020, doi: 10.1109/TIFS.2020.2980792.
- [44] Y. Wang, P. Bilinski, F. Bremond, and A. Dantcheva, "ImaGINator: Conditional Spatio-Temporal GAN for Video Generation," in 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA: IEEE, Mar. 2020, pp. 1149–1158. doi: 10.1109/WACV45572.2020.9093492.
- [45] C. Fu, Y. Hu, X. Wu, G. Wang, Q. Zhang, and R. He, "High-Fidelity Face Manipulation With Extreme Poses and Expressions," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 2218–2231, 2021, doi: 10.1109/TIFS.2021.3050065.
- [46] Y. Didi, "Jiggy: Magic dance gif maker (2020)," URL <https://apps.apple.com/us/app/jiggy-magic-dance-gif-maker/id1482608709>.
- [47] W. Liu, Z. Piao, J. Min, W. Luo, L. Ma, and S. Gao, "Liquid Warping GAN: A Unified Framework for Human Motion Imitation, Appearance Transfer and Novel View Synthesis," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 5904–5913. Accessed: Jun. 13, 2024. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/html/Liu_Liquid_Warping_GAN_A_Unified_Framework_for_Human_Motion_Imitation_ICCV_2019_paper.html
- [48] T. Xiao, J. Hong, and J. Ma, "ELEGANT: Exchanging Latent Encodings with GAN for Transferring Multiple Face Attributes," presented at the Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 168–184. Accessed: Jun. 13, 2024. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Taihong_Xiao_ELEGANT_Exchanging_Latent_ECCV_2018_paper.html
- [49] T. Li et al., "BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network," in Proceedings of the 26th ACM international conference on Multimedia, in MM '18. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 645–653. doi: 10.1145/3240508.3240618.
- [50] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "AttGAN: Facial Attribute Editing by Only Changing What You Want," IEEE Transactions on Image Processing, vol. 28, no. 11, pp. 5464–5478, Nov. 2019, doi: 10.1109/TIP.2019.2916751.
- [51] M. Liu et al., "STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3673–3682. Accessed: Jun. 14, 2024. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2019/html/Liu_STGAN_A_Unified_Selective_Transfer_Network_for_Arbitrary_Image_Attribute_CVPR_2019_paper.html
- [52] Y. Jo and J. Park, "SC-FEGAN: Face Editing Generative Adversarial Network With User's Sketch and Color," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1745–1753. Accessed: Jun. 15, 2024. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/html/Jo_SCFEGAN_Face_Editing_Generative_Adversarial_Network_With_Users_Sketch_and_ICCV_2019_paper.html
- [53] X. Nie, H. Ding, M. Qi, Y. Wang, and E. K. Wong, "URCA-GAN: UpSample Residual Channel-wise Attention Generative Adversarial Network for image-to-image translation," Neurocomputing, vol. 443, pp. 75–84, Jul. 2021, doi: 10.1016/j.neucom.2021.02.054.
- [54] "Photo-realistic face age progression/regression using a single generative adversarial network - ScienceDirect." Accessed: Jun. 14, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0925231219310926>
- [55] J. Guo and Y. Liu, "Attributes guided facial image completion," Neurocomputing, vol. 392, pp. 60–69, Jun. 2020, doi: 10.1016/j.neucom.2020.02.013.
- [56] M. Afifi, M. A. Brubaker, and M. S. Brown, "HistoGAN: Controlling Colors of GAN-Generated and Real Images via Color Histograms," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7941–7950. Accessed: Jun. 15, 2024. [Online]. Available: https://openaccess.thecvf.com/content_CVPR2021/html/Afifi_HistoGAN_Controlling_Colors_of_GAN_Generated_and_Real_Images_via_Color_CVPR_2021_paper.html
- [57] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4401–4410. Accessed: Jun. 05, 2024. [Online]. Available: http://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html
- [58] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE, Jun. 2020, pp. 8107–8116. doi: 10.1109/CVPR42600.2020.00813.

- [59] 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.
- [60] H. Nguyen and R. Derakhshani, *Eyebrow Recognition for Identifying Deepfake Videos*. 2020.
- [61] 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.
- [62] S. Agarwal, H. Farid, O. Fried, and M. Agrawala, "Detecting Deep-Fake Videos From Phoneme-Viseme Mismatches," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 660–661. Accessed: Jun. 18, 2024. [Online].: https://openaccess.thecvf.com/content_CVPRW_2020/html/w39/Agarwal_Detecting_Deep-Fake_Videos_From_Phoneme-Viseme_Mismatches_CVPRW_2020_paper.html
- [63] C.-Z. Yang, J. Ma, S. Wang, and A. W.-C. Liew, "Preventing DeepFake Attacks on Speaker Authentication by Dynamic Lip Movement Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1841–1854, 2021, doi: 10.1109/TIFS.2020.3045937.
- [64] 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.
- [65] K. Zhang, Y. Liang, J. Zhang, Z. Wang, and X. Li, "No One Can Escape: A General Approach to Detect Tampered and Generated Image," *IEEE Access*, vol. 7, pp. 129494–129503, 2019, doi: 10.1109/ACCESS.2019.2939812.
- [66] A. Khodabakhsh and C. Busch, "A Generalizable Deepfake Detector based on Neural Conditional Distribution Modelling".
- [67] "DeepFake Videos Detection Using Self-Supervised Decoupling Network." Accessed: Jun. 19, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9428368>
- [68] H.-S. Chen, M. Rouhsedaghat, H. Ghani, S. Hu, S. You, and C.-C. J. Kuo, "DefakeHop: A Light-Weight High-Performance Deepfake Detector," Mar. 11, 2021, arXiv: arXiv:2103.06929. Accessed: Jun. 19, 2024. [Online]. Available: <http://arxiv.org/abs/2103.06929>
- [69] J. Li, H. Xie, J. Li, Z. Wang, and Y. Zhang, "Frequency-Aware Discriminative Feature Learning Supervised by Single-Center Loss for Face Forgery Detection," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6458–6467. Accessed: Jun. 19, 2024. [Online]. Available: https://openaccess.thecvf.com/content_CVPR2021/html/Li_Frequency-Aware_Discriminative_Feature_Learning_Supervised_by_Single-Center_Loss_for_Face_CVPR_2021_paper.html
- [70] H. Liu et al., "Spatial-Phase Shallow Learning: Rethinking Face Forgery Detection in Frequency Domain," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA: IEEE, Jun. 2021, pp. 772–781. doi: 10.1109/CVPR46437.2021.00083.
- [71] I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, "Deepfake Video Detection through Optical Flow Based CNN," in 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South): IEEE, Oct. 2019, pp. 1205–1207. doi: 10.1109/ICCVW.2019.00152.
- [72] H. Qi et al., "DeepRhythm: Exposing DeepFakes with Attentional Visual Heartbeat Rhythms," in Proceedings of the 28th ACM International Conference on Multimedia, Seattle WA USA: ACM, Oct. 2020, pp. 4318–4327. doi: 10.1145/3394171.3413707.
- [73] S. Tariq, S. Lee, and S. S. Woo, "A Convolutional LSTM based Residual Network for Deepfake Video Detection," Sep. 16, 2020, arXiv: arXiv:2009.07480. Accessed: Jun. 20, 2024. [Online]. Available: <http://arxiv.org/abs/2009.07480>
- [74] M. Rana, M. Nobil, B. Murali, and A. Sung, "Deepfake Detection: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 1–1, Jan. 2022, doi: 10.1109/ACCESS.2022.3154404.
- [75] T. T. Nguyen et al., "Deep learning for deepfakes creation and detection: A survey," *Computer Vision and Image Understanding*, vol. 223, p. 103525, Oct. 2022, doi: 10.1016/j.cviu.2022.103525.
- [76] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," arXiv preprint arXiv:1811.00656, 2018.
- [77] 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.
- [78] Y.-X. Zhuang and C.-C. Hsu, "Detecting Generated Image Based on a Coupled Network with Two-Step Pairwise Learning," in 2019 IEEE International Conference on Image Processing (ICIP), Sep. 2019, pp. 3212–3216. doi: 10.1109/ICIP.2019.8803464.
- [79] C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, "Deep Fake Image Detection Based on Pairwise Learning," *Applied Sciences*, vol. 10, no. 1, Art. no. 1, Jan. 2020, doi: 10.3390/app10010370.
- [80] P. Kumar, M. Vatsa, and R. Singh, "Detecting Face2Face Facial Reenactment in Videos," in 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA: IEEE, Mar. 2020, pp. 2578–2586. doi: 10.1109/WACV45572.2020.9093628.
- [81] 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), Seattle, WA, USA: IEEE, Jun. 2020, pp. 5780–5789. doi: 10.1109/CVPR42600.2020.00582.
- [82] R. Wang et al., "FakeSpotter: A Simple yet Robust Baseline for Spotting AI-Synthesized Fake Faces," Jul. 16, 2020, arXiv: arXiv:1909.06122. Accessed: Jun. 28, 2024. [Online]. Available: <http://arxiv.org/abs/1909.06122>
- [83] X. Zhu, H. Wang, H. Fei, Z. Lei, and S. Z. Li, "Face Forgery Detection by 3D Decomposition," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA: IEEE, Jun. 2021, pp. 2928–2938. doi: 10.1109/CVPR46437.2021.00295.
- [84] Y. Luo, Y. Zhang, J. Yan, and W. Liu, "Generalizing Face Forgery Detection with High-frequency Features," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA: IEEE, Jun. 2021, pp. 16312–16321. doi: 10.1109/CVPR46437.2021.01605.
- [85] A. G. S and N. Thillaiarasu, "Investigation Of Comparison on Modified CNN Techniques to Classify Fake Face in Deepfake Videos," in 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Mar. 2022, pp. 702–707. doi: 10.1109/ICACCS54159.2022.9785092.
- [86] A. Raza, K. Munir, and M. Almutairi, "A Novel Deep Learning Approach for Deepfake Image Detection," *Applied Sciences*, vol. 12, no. 19, Art. no. 19, Jan. 2022, doi: 10.3390/app12199820.
- [87] A. Ismail, M. Elpeltagy, M. S. Zaki, and K. Eldahshan, "An integrated spatiotemporal-based methodology for deepfake detection," *Neural Comput & Applic*, vol. 34, no. 24, pp. 21777–21791, Dec. 2022, doi: 10.1007/s00521-022-07633-3.

- [88] J. B. Awotunde, R. G. Jimoh, A. L. Imoize, A. T. Abdulrazaq, C.-T. Li, and C.-C. Lee, "An Enhanced Deep Learning-Based DeepFake Video Detection and Classification System," *Electronics*, vol. 12, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/electronics12010087.
- [89] Y. PATEL et al., "An Improved Dense CNN Architecture for Deepfake Image Detection | IEEE Journals & Magazine | IEEE Xplore." Accessed: Jul. 02, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10057390>
- [90] A. Heidari, N. J. Navimipour, H. Dag, S. Talebi, and M. Unal, "A Novel Blockchain-Based Deepfake Detection Method Using Federated and Deep Learning Models," *Cogn Comput*, vol. 16, no. 3, pp. 1073–1091, May 2024, doi: 10.1007/s12559-024-1025-7.
- [91] A. Mehra, "Deepfake Detection using Capsule Networks with Long Short-Term Memory Networks".
- [92] I. Amerini and R. Caldelli, "Exploiting Prediction Error Inconsistencies through LSTM-based Classifiers to Detect Deepfake Videos," in *Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security*, in IH&MMSec '20. New York, NY, USA: Association for Computing Machinery, Jun. 2020, pp. 97–102. doi: 10.1145/3369412.3395070.
- [93] I. Masi, A. Killekar, R. M. Mascarenhas, S. P. Gurudatt, and W. AbdAlmageed, "Two-branch Recurrent Network for Isolating Deepfakes in Videos," Sep. 03, 2020, arXiv: arXiv:2008.03412. doi: 10.48550/arXiv.2008.03412.
- [94] Z. Sun, Y. Han, Z. Hua, N. Ruan, and W. Jia, "Improving the Efficiency and Robustness of Deepfakes Detection Through Precise Geometric Features," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3609–3618. Accessed: Jul. 08, 2024. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/html/Sun_Improving_the_Efficiency_and_Robustness_of_Deepfakes_Detection_Through_Precise_CVPR_2021_paper.html
- [95] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos," Jun. 17, 2019, arXiv: arXiv:1906.06876. Accessed: Feb. 29, 2024. [Online]. Available: <http://arxiv.org/abs/1906.06876>
- [96] S. A. Aduwala, M. Arigala, S. Desai, H. J. Quan, and M. Eirinaki, "Deepfake Detection using GAN Discriminators," in *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, Aug. 2021, pp. 69–77. doi: 10.1109/BigDataService52369.2021.00014.
- [97] Y. Huang, F. Juefei-Xu, Q. Guo, Y. Liu, and G. Pu, "FakeLocator: Robust Localization of GAN-Based Face Manipulations," *IEEE Trans. Inform. Forensic Secur.*, vol. 17, pp. 2657–2672, 2022, doi: 10.1109/TIFS.2022.3141262.
- [98] Preeti, M. Kumar, and H. K. Sharma, "A GAN-Based Model of Deepfake Detection in Social Media," *Procedia Computer Science*, vol. 218, pp. 2153–2162, Jan. 2023, doi: 10.1016/j.procs.2023.01.191.
- [99] P. Sharma, M. Kumar, and H. K. Sharma, "GAN-CNN Ensemble: A Robust Deepfake Model of Social Media Images Using Minimized Catastrophic Forgetting and Generative Replay Technique," *Procedia Computer Science*, vol. 235, pp. 948–960, 2024.
- [100] X. Li, K. Yu, S. Ji, Y. Wang, C. Wu, and H. Xue, "Fighting Against Deepfake: Patch&Pair Convolutional Neural Networks (PPCNN)." 2020, p. 89. doi: 10.1145/3366424.3382711.
- [101] I. Avatarify, "Avatarify: Ai face animator (2020)," URL <https://apps.apple.com/us/app/avatarify-ai-face-animator/id1512669147>.
- [102] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen, "Advancing High Fidelity Identity Swapping for Forgery Detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 5073–5082. doi: 10.1109/CVPR42600.2020.00512.
- [103] Z. Chen, L. Xie, S. Pang, Y. He, and B. Zhang, "MagDR: Mask-guided Detection and Reconstruction for Defending Deepfakes," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA: IEEE, Jun. 2021, pp. 9010–9019. doi: 10.1109/CVPR46437.2021.00890.
- [104] Alsaadi, Israa, Nuha Salim Mohammed, and Saja Salim Mohammed. "Optimizing Skin Disease Diagnosis using Metaheuristic Algorithms: A Comparative Study." *Iraqi Journal for Applied Sciences* 1.1 (2024): 72-80.
- [105] H. Liu et al., "Coherent adversarial deepfake video generation," *Signal Processing*, vol. 203, p. 108790, Feb. 2023, doi: 10.1016/j.sigpro.2022.108790.
- [106] KAMIL, IEHAB, and Mohanad Abdulsalam Younus AL-Askari. "NEW APPROACH TO PREDICTION OF MEMORY LEAK IN HPC HIGH-PERFORMANCE COMPUTING BY USING MPI (MESSAGE PASSING INTERFACE)." *Iraqi Journal for Applied Sciences* 1.1 (2024): 1-8.

BIOGRAPHIES OF AUTHORS

	<p>Israa Mishkhal was born in Iraq, in Baqubah. She obtained a bachelor's degree in computer science from Diyala University. She holds a master's degree in computer science from Ball State University (BSU) in the United States of America. She is a Ph.D. student at Universiti Sains Malaysia, Computer Science. She is a lecturer at College Science/Diyala University, Iraq (Diyala). She has many research papers in national and international conferences. email: israaadnan@uodiyala.edu.iq, israa_adnan85@student.usm.my</p> <p>Scopus®    </p>
	<p>Dr. Nibras Abdullah Born in Yemen .He is a distinguished academic and researcher, currently serving as a permanent faculty member at Hodeidah University, Yemen, and as an Assistant Professor (Senior Lecturer) at the School of Computer Sciences, Universiti Sains Malaysia, in Penang, Malaysia. He obtained a bachelor of Engineering from College of Engineering and Petroleum, Hadhramout University of Science and Technology, Yemen, 2003. He holds a master of Computer Science in Computer Sciences from Universiti Sains Malaysia, 2010 and a PhD in Computer Science: Specializing in Multimedia Network Protocols, National Advanced IPv6 Center of Excellence from Universiti Sains Malaysia, 2017. :Email: nibras@usm.my</p> <p>Scopus®    </p>