International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

A Comprehensive Survey on Deepfake Methods: Generation, Detection, and Applications

Battula Thirumaleshwari Devi[1], R Rajkumar[2]
[1]Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
maleshwari.devi@vit.ac.in
[2]Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
vitrajkumar@gmail.com

Abstract—Due to recent advancements in AI and deep learning, several methods and tools for multimedia transformation, known as deepfake, have emerged. A deepfake is a synthetic media where a person's resemblance is used to substitute their presence in an already-existing image or video. Deepfakes have both positive and negative implications. They can be used in politics to simulate events or speeches, in translation to provide natural-sounding translations, in education for virtual experiences, and in entertainment for realistic special effects. The emergence of deepfake face forgery on the internet has raised significant societal concerns. As a result, detecting these forgeries has become an emerging field of research, and many deepfake detection methods have been proposed. This paper has introduced deepfakes and explained the different types of deepfakes that exist. It also explains a summary of various deep fake generation techniques, both traditional and AI detection techniques. Datasets used for deepfake-generating that are freely accessible are emphasized. To further advance the deepfake research field, we aim to provide relevant research findings, identify existing gaps, and propose emerging trends for future study.

Keywords-Deep learning, Artificial intelligence, Deep fake, Multimedia, Deepfake Detection, Deepfake generation.

I. INTRODUCTION

Deepfake refers to photorealistic videos or images created with the support of Artificial Intelligence (AI) approaches. Creation of false content by altering the face of a source individual and superimposing it onto a picture or video of an individual else, known as the target. The result is a manipulated video or image that appears to be of the targeted individual, but actually, the face of the source is being used. "Deepfake" was coined after a Reddit account named "Deepfake" claimed to have developed a machine-learning technique for substituting popular characters into explicit content. This technique uses artificial intelligence to manipulate media and create convincing yet fraudulent content. Understanding the possible risks associated with Deepfake technology is crucial and to exercise caution when encountering media online [1].

Recently, fake videos and images circulating online have become a public issue as they can easily exploit individuals. The manipulated media outlets may have detrimental effects on people's reputations, disseminate misleading information, and even endanger their lives. It is important to remain vigilant and cautious when encountering media on the internet and verify the content's authenticity before sharing it with others. Additionally, this method is employed to promote hoaxes, fraud, and fake news [2]. Recently, researchers have become increasingly concerned about Deepfake technology and are dedicating their efforts to gain insights into its implications. With its potential to create fraudulent and misleading media, Deepfake has become a serious threat to individuals and society. Researchers are working tirelessly to uncover the inner workings of this technology and develop solutions to detect and prevent its harmful effects. We need to stay informed about the latest developments in this field and take necessary precautions to protect ourselves from the dangers of Deepfake technology [3].

Two neural networks were employed to create fraudulent videos: a discriminative and generative network, which utilized the FaceSwap approach. The discriminative network determines if the freshly created pictures are legitimate, while the generative network creates fake images using an encoder and decoder. Ian Goodfellow first suggested the idea of Generative Adversarial Networks (GANs), which are the merging of these two networks. [4]. Generative Adversarial Networks mimic a person's emotions and facial expressions and replace them with those of a different person. Deepfake technology has been particularly harmful to political figures, public figures, and celebrities. The ability to create convincing imitations of these individuals has made them the main targets of Deepfake creators. In addition, Deepfake technology has been used to spread false messages to world leaders, posing a significant threat to world peace. The potential for this technology to manipulate media and create fraudulent content is a serious

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

concern for individuals and society. It is crucial to stay informed about the latest developments in this field and take necessary precautions to prevent the harmful effects of Deepfake technology. Providing fake maps could mislead military personnel and cause damage [5].

This technology has many stimulating potentials despite its possible drawbacks. For example, those who have lost their voice can get it back with the use of this cutting-edge technology. We also know that fake news travels farther and more quickly than the truth. Unfortunately, recovering from its effects is a difficult task [6]. To fully understand Deepfakes, we must delve deeply into the subject matter. This includes understanding what Deepfakes are, how they are created, and how to detect them [7]. Since this field is relatively new and was only introduced in 2017, limited resources are available. However, several recent studies have been introduced to address the issue of social media misinformation related to Deepfakes. Although the creation of fake content is well-known, Deepfake takes it a step further by applying AI and ML to modify the initial information into almost real-looking fraudulent material [8].

Detection of Deepfakes through artificial intelligence has two primary goals: to identify and stop the spread of modified multimedia material produced by deep learning methods and to prevent deepfakes from occurring in the first place [9]. Detection of synthetic material, such as images, text, audio, and videos manipulated to mislead viewers, is one of the goals of AI-driven deepfake detection, which seeks to protect the honesty of digital media by properly recognizing instances of such content. Helping to battle disinformation, protecting people's reputations, and maintaining confidence in digital media platforms may be achieved by developing strong and efficient techniques to differentiate between authentic and altered material. These approaches will be developed using sophisticated algorithms and models [10]. Deepfake can be used for various things, including creating fake celebrity pornography, disseminating false information, impersonating politicians, committing financial fraud, etc. Face swapping has been a well-known technique in the film industry for creating fake voices or videos that meet their needs [11]. However, this process requires a lot of time and a certain level of expertise. Thanks to deep learning techniques, anyone with solid computer knowledge and access to a high-configuration GPU computer can create convincing fake videos or images. It's important to be aware of the potential misuse of such technology and to remain vigilant when consuming media [12].

manuscripts must be in English. These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts. Please follow them and if you have any questions, direct them to the production editor in charge of your proceedings at Conference Publishing Services (CPS): Phone +1 (714) 821-8380 or Fax +1 (714) 761-1784.

II. RELATED WORK

Detecting manipulated videos has become a major challenge in today's era where the creation and spread of such videos is effortless. In response to this challenge, a recent research paper [13], presents Deepfake Stack, a deep ensemble learning technique that outperforms other classifiers, achieving an accuracy of 99.65% and AUROC of 1.0. The authors of the paper introduce a lightweight 3D convolutional neural network for Deepfake detection. The network utilizes a channel transformation module to extract features with fewer parameters and a spatial-temporal module to fuse the spatial features in the time dimension. The experimental results demonstrate that the proposed approach surpasses other state-of-the-art Deepfake detection methods, making it a promising solution for detecting manipulated videos. Overall, the proposed method provides an effective and practical solution to the critical issue of detecting manipulated videos. Its high accuracy and AUROC values, combined with its lightweight architecture, make it an attractive option for practical applications. The results of the experiment provide strong evidence of the effectiveness of this approach, and we are confident that this method can be used to detect manipulated videos with high accuracy. According to a recent study [14], a Convolutional Vision Transformer (CNN + ViT) was proposed by researchers to detect Deepfake videos. The model produced remarkable results, boasting 91.5% accuracy, an AUC value of 0.91, and a loss value of 0.32 on the DeepFake Detection Challenge Dataset (DFDC). By integrating a CNN module into the ViT architecture, the model was able to achieve strong results on the DFDC dataset. The authors of reference [16] present a temporal-aware pipeline that enables the automatic identification of manipulated videos, also known as deepfakes. To achieve this, the pipeline employs a convolutional neural network (CNN) to extract frame-level features, and a recurrent neural network (RNN) to classify the videos as either manipulated or unmanipulated. The results of the study demonstrate that the pipeline's performance is competitive. The paper [3] presents a novel and effective approach for real-time facial reenactment of a monocular target video sequence. The proposed method employs non-rigid model-based bundling for facial expression tracking and deformation transfer to achieve highly realistic reenactment results. The output is then photo-realistically re-rendered to seamlessly blend with the real-world illumination, resulting in exceptional visual quality.

The authors of [17] have approached the issue of detecting deepfake images by formulating it as a hypothesis testing problem. They have employed a robust statistical analysis of GANs to effectively limit the probability of errors. Additionally, they have established a correlation between error probability and epidemic thresholds for spreading processes in networks. In a recent study, the creators of [18] shared a collection of Deepfake videos that are accessible to the public. Their findings revealed that even the most sophisticated facial recognition systems are

vulnerable to Deepfakes. Moreover, they observed that the audio-visual technique used to detect lip-sync inconsistencies was ineffective in distinguishing Deepfakes. The top-performing method resulted in an 8.97% equal error rate for high-quality Deepfakes. In [19], proposes a new technique for detecting fake face videos generated with neural networks by detecting eye blinking. The proposed method has been tested against benchmarks of eye-blinking detection datasets and has shown considerable promise in detecting DeepFake videos. The methods for detecting edited images generated by artificial intelligence were explored in a study published as [20]. The authors analyzed the artifacts left behind by facial editing techniques and showed that even slight visual distortions can effectively reveal manipulation. These techniques are not only easy to implement but also allow for quick adjustments, achieving impressive AUC values of up to 0.866. Within [21], an innovative approach is proposed that employs deep learning to differentiate between AI-generated fake videos and legitimate videos. This approach utilizes convolutional neural networks to identify specific artifacts that result from affine face warping. The method is highly efficient and resourceful, outperforming other available techniques in terms of robustness.

The DeepLabv3+ model proposed by the authors in [1] is a robust solution for improving semantic segmentation. By utilizing spatial pyramid pooling and encoder-decoder structure, the model has attained impressive test set performance on both PASCAL and Cityscapes datasets, achieving a remarkable 89% and 82.1%, respectively, without requiring any post-processing. FSGAN, proposed by [2], is a state-of-the-art approach for subject-agnostic face swapping and reenactment. The technique leverages a recurrent neural network, incorporating a face completion network and a face blending network to produce exceptional results that surpass those of existing systems. With FSGAN, users can achieve high-quality face swapping and reenactment, regardless of the subject, with a level of accuracy that was previously unattainable. This technology is poised to enhance the field of facial recognition and holds immense potential for a wide range of applications in the business and academic settings. In a recent paper [3], a novel technique for achieving real-time facial reenactment of a monocular target video sequence was introduced. This method involves utilizing non-rigid model-based bundling to track facial expressions and transfer deformation, resulting in highly accurate reenactment results. To enhance the realism of the output, the system is designed to re-render the results in a photo-realistic manner that seamlessly blends with real-world illumination. The result is a visually stunning and remarkably lifelike facial reenactment that can be achieved in real-time. In this groundbreaking research paper, a novel method for creating high-resolution, photo-realistic images is presented in [4]. The approach employs conditional generative adversarial networks, resulting in visually stunning and easily customizable outcomes. Along with object manipulations, the technique is capable of generating diverse results, rendering it highly adaptable. The system's end product boasts an impressive level of realism and aesthetic appeal, making it an ideal solution for a wide range of image generation applications. FReeNet is an innovative facial reenactment framework designed by [5], utilizing advanced techniques to transfer facial expressions from any source face to a specific target face. The primary goal is to generate remarkable photo-realistic and expression-like faces that resemble the target face while maintaining the original facial expressions of the source face. The framework comprises two significant components: the Unified Landmark Converter (ULC) and the geometry-Aware Generator (GAG). The ULC is responsible for converting the landmarks of the source face to match those of the target face, ensuring that the generated face is perfectly aligned with the target face. The GAG is then utilized to produce a face that is both photorealistic and expressive, resulting in a remarkable output.

In the publication referenced as [6], the authors present an innovative approach to face swapping that utilizes Convolutional Neural Networks (CNNs) and style transfer. Their method introduces a distinctive loss function that elevates the authenticity of the resulting images and enables seamless face swapping in real-time, without the need for user input. The technique employs an encoder-decoder structure and incorporates perceptual and style losses to generate visually appealing faces. [7] this remarkable paper presents the Labelled Faces in the Wild database, an extensive resource that provides an outstanding opportunity to investigate face recognition in real-world settings. The database includes labelled photographs of faces that exhibit natural variation in numerous factors such as pose, lighting, race, accessories, occlusions, and background. Furthermore, the database features baseline outcomes and parallel databases that can effectively evaluate the performance of face recognition algorithms. According to a research paper authored by [8], the Super-Resolution GAN (SRGAN) is an advanced generative adversarial network that has been specially developed to enhance image resolution. This technique employs a deep residual network and a perceptual loss function to restore photo-realistic textures from images that have been heavily down sampled. The researchers conducted extensive mean-opinion-score (MOS) testing, which demonstrated that SRGAN significantly improves the perceptual quality of generated images. The authors of this paper introduce StarGAN [9], which is a comprehensive generative adversarial network designed to enable image-to-image translation across multiple domains. The model was tested and evaluated on the CelebA dataset, and it effectively demonstrates that a single generator network can produce facial expression label images from the RaFD dataset. This approach offers a unified and efficient solution for multi-domain image-to-image translation, which has been a challenging problem in the field. The StarGAN model leverages

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

adversarial training to learn to generate high-quality images that maintain the structure and content of the input image while transforming it into a target domain. The evaluation of the model on the CelebA dataset shows that it can perform well on a diverse range of tasks, including facial attribute manipulation, face aging, and gender conversion. The results demonstrate the effectiveness of the model and its ability to generate high-quality images with rich and varied content. Overall, the StarGAN model represents a significant advancement in the field of generative adversarial networks and offers a promising direction for future research in this area. The proposed approach for identity-preserving face synthesis is the FaceID-GAN [10] model. Its three-player GAN and symmetry allow for generating faces from any viewpoint while maintaining the identity. The approach shows improvement over previous techniques, and its potential for further research and development is promising.

TABLE1. OVERVIEW OF EXISTING DEEPFAKE GENERATION METHODS

S. No	Reference	Deepfake Type	Technique	Generated output
1	[1][2][11] [12][13][6]	FS	Autoencoder, loss functions, and statistical attention mechanism Autoregressive model, various loss functions, GAN, and CNN	Image and video
2	[14][15][16] [17][5][18] [19][20][21] [22][10][23] [9][24][25] [26][27][28]	F2F	Autoencoder, CNN, and various loss functions, GAN, tertiary processing methods, Autoregressive model, statistical attention mechanism, GAN, RNN, autoencoder, GAN, and various loss functions	Image and video
3	[29][30][31] [32][33][34] [35][36]	FAM	GAN, various loss functions, statistical attention mechanism, autoencoder, and RNN	Image and video
4	[37][38][39] [40]	EFS	GAN, loss functions, and statistical attention mechanism	Image and video
5	[41][42][43] [44][45][46] [47][48][49] [50][51][52]	MR	GAN and loss functions, statistical attention mechanism, tertiary processing methods, CNN, and RNN	Image and video
6	[53][54]	NT	Autoencoder, RNN, tertiary processing methods, and statistical attention mechanism	Image and video

*Note: GAN (Generative Adversarial Network), RNN (Recurrent Neural Network), CNN (Convolutional Neural Network), FS(FaceSwap), F2F(Face2Face), NT (Neural Textures).

Figure 1 depicts the working principle of different deepfake generation methods, and Figure 2 illustrates the detection techniques for different types of deepfakes.

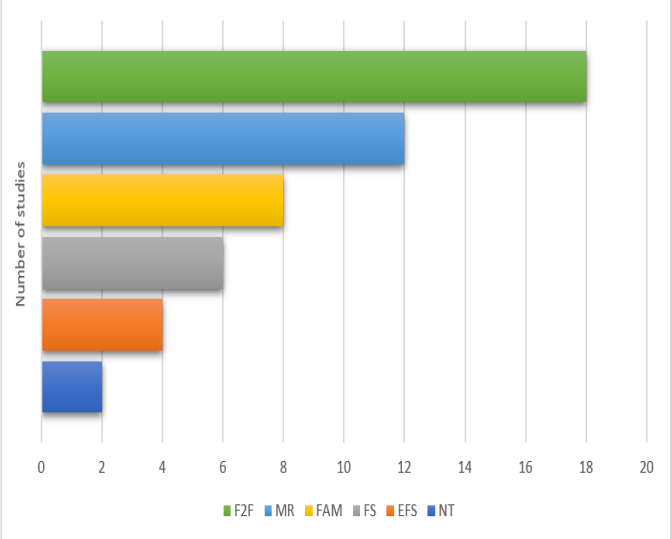

Figure 1. The basic architecture of convolutional neural network [72].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

TABLEII. OVERVIEW OF EXISTING DEEPFAKE DETECTION METHODS

S. No	Reference	Deepfake Type	Method	Dataset
1	[1][56][3][56][4][5] [6][7][8][9][10] [11][12][13][14][15] [16][17][18][20][75] [76][23][24][25][26] [81][28][29][30][1]	DMF	MIL, SVM, DA, KNN, EM, ADB, CRA, CNN, EM, RNN, MLP, LR, ETH, STAT	CELEB-DF, FF, DFDC, FF+, CELEB-A, DF-TIMIT
2	[32][33][34][35][36] [37][38][9][40][41] [42][43] [44][45][46] [47][102][49][50][105] [52][53][54][55][56] [57][58][59][60][1] [62][63][63][64][65] [66][67][8][69][70][71] [126][73][74][75][76] [131][132][133][80] [81][82][137] [138][85][140][87][88]	FM	SVM, LR, K-NN, MTCNN, RNN, MLP, CNN, RF, NB, LR, K-NN, DT, HMN, XGB, ADB	CELEB-A, CELEB-DF, FS, FF++, FF, DFD, DFDC, DF-1.0, DF-TIMIT, UADFV, FE, FFW, MANFA, SMFW,

*Note: DMF (Digital Media Forensics), FM (Face Manipulation), CNN: (Convolutional Neural Network), RNN: (Recurrent Neural Network), RCNN: (Regional Convolutional Neural Network), MTCNN: (Multi-task Cascaded CNN, MSCNN: multi-scale Temporal CNN, MIL (Multiple Instance Learning), SVM (Support Vector Machine), LR (Logistic Regression), MLP (Multilayer Perceptron), ADB (Adaptive Boosting), XGB (eXtreme Gradient Boosting), DT (Decision Tree), NB (Naive Bayes), RF (Random Forest), k-MN: K means clustering, KNN: K-Nearest Neighbour, DA: Discriminant Analysis, (EM: Expectation Maximization, CRA: Co-relation Analysis, ETH: Ethereum Blockchain.

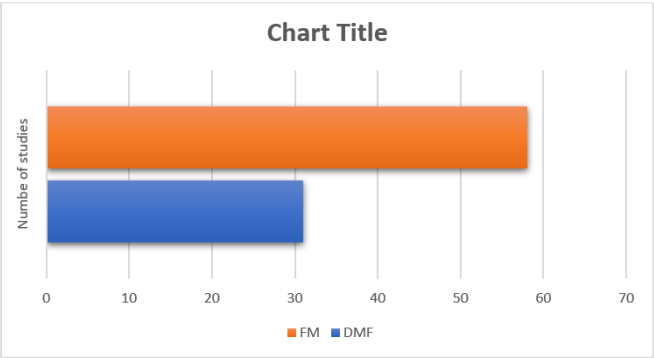

Figure 2. Study distribution with respect to deep fake detection methods.

III. DEEPFAKE TYPES

A. Full Face Synthesis

The full-face synthesis process generates a realistic and complete image of a human face from scratch. This task is challenging because it requires capturing human faces' subtle variations and expressions while avoiding artifacts and distortions. Face synthesis creates a synthetic personality that is incredibly lifelike by first learning the hidden representation of a face dataset. Since the face synthesis approach generates its persona without a target subject, it does not exist in reality. Although this method helps the modeling and gaming sectors, attackers might use it to fabricate identities for illicit purposes [163]. Some full-face synthesis applications are Virtual and augmented reality, Film and video production, Advertising and marketing, and Security.

There are two primary methods for generating a complete face:

1) GAN (Generative Adversarial Network): A technique for machine learning called GAN can produce realistic visuals. GANs are trained on a dataset of real human faces in the context of face synthesis. The GAN gains the ability to create fresh faces that resemble those in the training dataset [164].

2) Variational Auto Encoders (VAEs): One machine learning algorithm that can be used to create realistic images is called VAEs. These algorithms are trained on a dataset of actual human faces. The VAEs learn to represent these faces as a latent distribution, which can then be sampled to generate new faces. In simpler terms, VAEs can generate new images of human faces by learning from real examples [165].

B. Reenactment

Facial reenactment is a process that goes beyond face synthesis. Instead of generating a new face, it involves disseminating body language or facial emotions from one individual to another. In other words, facial reenactment allows for the replication of the movements and expressions of one person onto another. Facial reenactment has been a popular technique for some time, predating the rise of deepfakes. Traditional approaches to facial reenactment rely on computer graphics to achieve the desired results [166]. These techniques involve the manipulation of pre-existing images or video footage to create a reenactment of the facial expressions or body movements of one person onto another.

The studies of various reenactment techniques will be outlined, including neural textures enable the transmission of facial expressions, while Face2Face and Puppet Master technologies allow for realistic facial and body movements.

1) Neural textures: Creating realistic visuals often requires the use of parametric vectors through feature maps, texture maps, or neural textures. In a study published in [74], authors presented an end-to-end delayed neural rendering network that blends learnable neural textures with conventional computer graphics expertise. The network is designed based on convolutional encoder-decoders. The authors generated a UV map that matched the expression of the source and sampled the neural textures of the target. This UV map was then fed into the neural renderer and backdrop image to carry out the synthesis process and produce the final reenactment result [167]. However, it's important to note that the quality of the output can be easily affected by the geometry proxy used.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

2) Face2Face: Face2Face is a widely used method for face reenactment that involves projecting a source's expressions onto a target. It's crucial to remember, nevertheless, that this technique could be exploited by attackers to manipulate or synthesize expressions on the target's face, making them appear to say or do things that they would never actually say or do [168]. The 'Obama deepfake video' is a well-known example of the misuse of this reenactment method. The main dataset for model training in a number of research has been Barack Obama's making it easier for attackers to create deepfake videos of him. To create the 'Obama deepfake video,' the creators extracted audio elements identified in the original video using Mel-frequency Cepstral Coefficients (MFCC), the researchers converted the target's mouth movements into vectors and applied Principal Component Analysis (PCA) on the frames to represent the mouth shape. After that, they trained a Long Short-Term Memory (LSTM) network to map the MFCC audio coefficients to the PCA mouth-shape coefficients. This approach helps to better capture the nuances of spoken words and improve the accuracy of speech recognition systems. [46]. In 2017, a fully trainable reenactment network capable of supporting lip synthesis based on text input was created by Kumar and his colleagues. A dataset similar to the one in the earlier study was employed in their investigation. The network was composed of three primary modules: 1) an LSTM network to convert audio to mouth key points, 2) a Pix2Pix network based on UNet architecture to synthesize target video based on the mask and mouth key points, and 3) a text-to-speech network named CharWav. This method can assist in creating more lifelike lip motions for virtual actors and enhance the quality of synthesized video output [169].

3) Motion Renactant (Puppet Master): It is acknowledged that this technique achieves in order to achieve a high degree of photorealistic motion transfer, it is crucial to preserve the original appearance of the body when transferring its motion or position from a source to a destination. However, pixel-to-pixel misalignment is a typical problem caused by differences between the source and destination. The authors [170], of this text proposed a solution that utilizes a deformable skip-connection with the nearest neighbor loss. The idea behind this approach is to break down the global information of the source into a local affine transformation set based on the target pose and then deform the source's feature map. They then applied a common skip connection to transfer the transformed tensor and fused it with other corresponding tensors in the decoder to generate the synthetic output.

C. Facial Attribute Manipulation

Facial attribute manipulation refers to modifying certain features of the face, such as gender, age, skin tone, hairstyles, wrinkles, and eye color. This method can change someone's appearance. based on the pre-set conditions. Facial attribute manipulation can be achieved using various techniques, including 1. Deep neural networks 2. Generative Adversarial Networks (GANs) 3. Autoencoders 4. Conditional GANs 5. StyleGAN 6. StarGAN 7. CycleGAN 8. Progressive Growing GANs (PGGANs). Facial attribute manipulation can also be achieved through manual editing of the appearance of a face [171]. This can involve adding or removing wrinkles, changing the shape of the nose or lips, or altering hair color. However, manual editing is a time-consuming and labor-intensive process that requires significant skill and expertise. A neural network architecture called StarGAN is used to manipulate face images across several domains. Transforming an input image from one domain to another while maintaining the underlying face features is its fundamental goal. StarGAN [172] has several advantages over previous studies, one of which is its multi-domain training capability. This is accomplished by the use of a mask vector approach, which enables the model to recognise and work with just the pertinent face features during translation. Conversely, previous studies concentrated on translating styles across two domains, which limited their capacity to control certain face features. Because of its multi-domain training capacity, StarGAN is a useful tool for a variety of face modification applications, such as virtual cosmetics, facial emotion transfer, and facial recognition.

D. Face Swapping

Another popular technique used to manipulate facial attributes is face swapping. Face swapping is the process of changing a person's face in a picture or video while maintaining the other parts of the original source. This technique can be used for various purposes, such as entertainment, social media, and even forensic investigations [26]. A Python program called FaceSwap was developed by students from Warsaw University of Technology. The program uses picture blending, Gauss-Newton optimization, and a face-alignment technique based on Deep Alignment Network to swap faces. First the algorithm detects a given input image's face region and landmarks. The texture coordinates are created by projecting the 3D model vertices onto the picture space after fitting it to landmarks. The algorithm advances in five phases after the first procedure.

- Identifying the facial landmarks and the face area.
- Fitting the 3D models to match the identified landmarks.
- Rendering the 3D model.
- The process of combining the rendered and camera images involves the use of color correction and alpha-blending techniques.
- Presenting the final image.

In 2017, [32] Korshunova and colleagues proposed a fast face-swapping technique to change person A's identity to person B while maintaining the same facial gesture, head location, and light intensity. Contrary typical style transfer techniques, this approach uses a person's identity as the "style" and preserves the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

remaining features as the "content." In order to achieve the face-swapping effect, the authors employed a modified multiscale convolutional neural network (CNN). This network had a content loss, style loss, and light loss functions included. The resulting face was aligned using an affine transformation that made use of 68 facial key points, while the backdrop was stitched together with a segmentation mask. The neural network's ability to swap with several target individuals is due to its one-to-many face swap method, which has been trained to comprehend person A's content. However, for fine-tuning, the training process requires a large number of single-person image datasets, which may not be feasible for general applications.

III. GENERATION TECHNIQUES

A. Deepfake Generation Techniques

The major drawbacks of conventional forgery production techniques are addressed by the innovative media manipulation method known as "deep fake generation." Deepfake creation, in contrast to conventional techniques, minimizes manipulation traces or fingerprints that are commonly used for forgery detection, such as biometric or compression artifact inconsistencies. This is achieved through advanced machine learning algorithms that can synthesize media that closely resembles real content. However, this technology also poses a significant threat as it can be used to create convincing fake media, which can be used to manipulate public opinion, spread fake news, or damage reputations. It is important to be aware of this technology's existence and take measures to prevent its misuse [173]. Deepfake technology utilizes a deep neural network that has been trained to recognize patterns in incoming data and use that information to generate highly realistic videos. To The three models that are commonly used to create deepfakes are the Autoregressive model [175], the Autoencoder [176], and the Generative Adversarial Network (GAN). achieve this, the network aims to learn the segmentation map or latent representation. Unlike traditional manipulation media, identifying deepfakes is becoming more challenging as the differences between genuine and fabricated data are minimal. This is because they can be used to create highly convincing and realistic content that is difficult to distinguish from genuine media. Therefore, developing robust mechanisms for detecting deepfakes is essential to prevent their misuse and minimize their potential impact [174]. The three general deepfake creation models are i. Autoregressive model [175], ii. Autoencoder [176], and iii. Generative Adversarial Network (GAN) [4].

1) Autoregressive Model: Autoregressive models are designed to focus on the natural distribution of images rather than their latent representation. Essentially, certain models employ a sequential review and pixel-by-pixel predictions to ascertain the conditional distribution of every pixel and its relationship to its predecessors in order to produce high-quality images. While there are two instances of this technique—Pixel-RNN and Pixel-CNN—the assessment procedure might take a while [177].

2) Autoencoder: An artificial neural network called an autoencoder is used to learn data representation unsupervised. It is made up of an encoder and decoder network that are connected to one another. The input is changed by the encoder into a latent, hidden representation, which the decoder then uses to reconstruct the original data representation. Making an output that closely mimics the input is the goal. Deepfakes are developed using the variational autoencoder (VAE), which is an essential component.

The foundation of an autoencoder is network training to recognize the essential features of the input while ignoring any unrelated noise. While similar to a conventional autoencoder, the variational autoencoder (VAE) differs because The Variational Autoencoder (VAE) is known for its effectiveness in restoring complex input data, which can be attributed to the strong assumption made in the probability distribution of latent variables. However, it's worth noting that the VAE has a higher probability of generating blurry or unclear outputs, even though it can generate new output data after training through sampling the distribution [179].

3) Generative Adversarial Network (GAN): A generator network and a discriminator network are the two neural networks that make up a Generative Adversarial Network (GAN). The goal of the generator network is to fool the discriminator by creating a new synthetic output depending on the data distribution of the input. The discriminator network, on the other hand, seeks to precisely determine if the output sample is authentic or fraudulent. Both networks use backpropagation to continuously optimize until they achieve an equilibrium state where it is impossible to tell the difference between the actual and false data. GAN can do alterations that are challenging to perform with conventional forgeries generation techniques, such as style transfer and picture restoration.[180].

Deepfake generation tools are software applications that use cutting-edge machine learning algorithms to create manipulated media content, such as videos and images, that appear real but fake. These tools have raised concerns about the potential misuse of AI-powered technology and its impact on public trust and safety. While deepfake technology can be used for entertainment, it has also been used to spread disinformation and fake news, leading to serious consequences for individuals and society. Therefore, developing effective strategies and policies to regulate the use of deepfake generation tools and prevent their misuse is crucial. Table 3 illustrates the various deepfake generation tools, and their descriptions are presented.

TABLE III. OVERVIEW OF VARIOUS APPS AND TOOLS USED FOR DEEPFAKE GENERATION

S. No	Tool	Platform	Description	Link
1	FaceApp	Mobile app	Uses AI to improve the image quality. It provides a wide range of filters and effects, including some that can be used to create deepfakes.	https://www.faceapp.com/
2	FaceSwap	TensorFlow	It is an open-source deepfake tool that can be used to swap faces in images and videos	https://faceswap.dev/
3	AutoFaceSwap	Desktop application	Upload the image or video in which you want to swap faces, and select the two faces that you want to swap. AutoFaceSwap will then automatically train a deep-learning model on the two faces and swap them in the image or video.	https://apps.microsoft.com/store/detail/auto-face-swap/9NBLGGH3M5NQ?hl=en-us&gl=us
4	e	TensorFlow	Open-source deepfake tool that can create realistic-looking deepfakes of faces in images and videos.	https://github.com/iperov/DeepFaceLab
5	FaceSwap-GAN	TensorFlow	Uses generative adversarial networks (GANs) to develop realistic-looking deepfakes of faces in images and videos	https://github.com/shaoanlu/faceswap-GAN
6	StarGAN	PyTorch	Uses GANs to learn mappings among multiple domains of facial images, videos, and actions	https://fsjournal.cpu.edu.tw/content/vol20.no.1/FSJ2021No1.pdf
7	StarGAN-V2	PyTorch	Extension of StarGAN developed by researchers at Naver for generating high-quality realistic images of human faces from a single image	https://github.com/clovaai/stargan-v2
8	FSGAN	PyTorch	Uses GANs to swap faces in images and videos without requiring training on the specific facts involved	https://github.com/YuvalNirkin/fsgan
9	StyleGAN	TensorFlow	Developed by researchers at NVIDIA for generating high-quality realistic images of human faces	https://github.com/NVlabs/stylegan
10	StyleGAN2	TensorFlow	Extension version of StyleGAN	https://github.com/NVlabs/stylegan2
11	StyleGAN2-ADA	PyTorch	A successor to the StyleGAN2 model, it incorporates a new technique called Adaptive Discriminator Augmentation (ADA) to enhance the generated image quality even for small dataset	https://github.com/NVlabs/stylegan2-ada
12	DiscoFaceGAN	TensorFlow	Based on GAN and expertise in facial expression transfer	https://github.com/microsoft/DiscoFaceGAN
13	ZAO	Mobile application	Allows users to swap their faces onto videos of celebrities, politicians, and other public figures. Momo, a Chinese social media company, developed it	https://zao.en.uptodown.com/android
13	DFaker	TensorFlow	Creates realistic and engaging facial expressions for avatars and virtual characters	https://github.com/dfaker/df
14	DeepFake_tf	TensorFlow	Similar to DFaker	https://github.com/StromWine/DeepFake_tf
15	DeepFakesWeb	Web application	Uses AI and DL to generate face swap videos	https://deepfakesweb.com/
16	Jiggy	Mobile application	Creates videos and GIFs of people dancing	https://play.google.com/store/apps/details?id=com.trainerize.jiggy
17	Avatarify	Mobile and web application	Animate any image with your facial movements	https://avatarify.ai/

18	Reface	Mobile and web application	Swap faces in the GIFs, memes, and videos with just one photo.	https://reface.ai/
19	MachineTube	Mobile application	Face swap in both image and video	https://appsgeyser.io/12957453/MachineTube
20	FewShotFace translation	TensorFlow	GAN-based model for face swap	https://github.com/shaoanlu/fewshot-face-translation-GAN

IV. DEEPFAKE DETECTION TECHNIQUES

Deepfake detection techniques refer to a broad range of methods used to identify and authenticate videos or images created using deep learning algorithms or AI tools. These techniques are primarily based on analyzing the discrepancies between the manipulated content and the original authentic content [55] [56] [57] [58]. Deepfakes are created by using AI algorithms that train on images and videos of a person's face and then generate realistic content that can be used to create an appearance that someone is acting or speaking something they have never actually done or said. Deepfakes are a growing concern due to their potential to cause significant harm, such as spreading fake news or propaganda, manipulating public opinion, and damaging an individual's reputation [59]. Deepfake detection techniques help combat these harms by utilizing various methods, such as analyzing facial expressions speech patterns, and identifying inconsistencies in the manipulated content. Effective deepfake detection requires a multidisciplinary approach involving computer vision, machine learning, and forensic analysis expertise. As deepfake technology continues to evolve, developing and implementing robust deepfake detection techniques is crucial to mitigate its harmful effects on individuals and society [60].

This section aims to provide a literature review of the latest works that utilize DL-based approaches for detecting deepfakes. With the increasing sophistication of deepfake technology, there has been a growing need for developing effective detection techniques. Deep learning algorithms have shown significant promise in this area, with researchers exploring various approaches to detect manipulated media, such as videos and images. Recent studies have explored a range of DL-based approaches, including CNNs [61] [62] [63] [64], GANs [65] [66], and RNNs [67] [68]. These approaches have demonstrated encouraging findings in identifying deepfakes, with some achieving high accuracy rates.

A. Convolutional Neural Networks (CNNs)

CNNs are a type of artificial neural networks exceptionally adept at processing image, speech, or audio signals. These networks use a layered architecture to systematically analyze the input data and automatically extract relevant features, making them ideal for object detection, facial recognition, and speech synthesis applications. Three different types of layers make up a convolutional neural network (CNN): pooling, convolutional, and fully-connected (FC). The pooling or convolutional layers come after the convolutional layer, which is the first layer. The fully-connected layer, which is the last layer of the CNN, is shown in figure 3 [69] [70]. Larger regions of the image are recognised by the CNN as it processes the visual input, making it more complex. While the latter layers detect the item's bigger components or forms until it locates the intended object, the early levels concentrate on fundamental properties like colour and edges. CNNs are powerful models that can learn complex representations of the input data and achieve state-of-the-art attain cutting-edge performance across a variety of activities [71]. They have revolutionized the fields of natural language processing, speech recognition, and computer vision and are widely used in industry and academia. By leveraging the power of deep learning, CNNs have opened up new frontiers in artificial intelligence and are driving innovation in many domains.

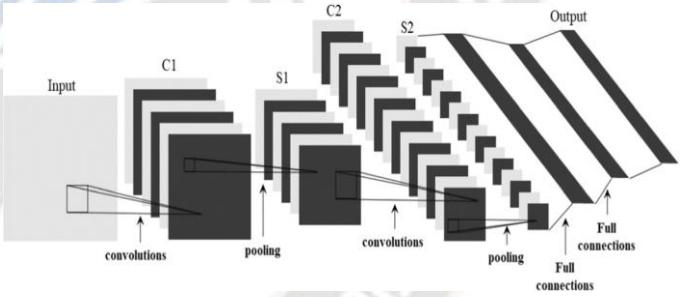

Figure3. The basic architecture of convolutional neural network [72].

1) Convolutional Layer: Convolutional layer is the main building block of this type of neural network. It requires a feature map, a filter, and input data to perform most of its calculations. Usually, the input data is made up of a 3D matrix of color pixels with height, width, and depth that correspond to RGB values in a picture. A feature detector, also known as a filter or kernel, scans the image's receptive fields to determine whether a feature is present. This process is called convolution, as shown in figure 4. A 2-D array of weights is used as a feature detector to find features in an image [73][74]. The receptive field's dimensions, which are typically 3 by 3 matrices but can change, are determined by the filter size. After applying the filter to a certain area of the picture, the dot product between the input and filter pixels is calculated. The dot product is loaded into an output array known as a feature map, activation map, or convolved feature after this procedure is repeated. Once the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

kernel has swept across the whole picture, the filter moves one step at a time. The picture starts to identify bigger components or forms of the item as it moves through the layers of the CNN, and eventually it recognises the desired thing. The convolutional layer is where most of the computation occurs in a CNN, and it is the core building block of the network [75].

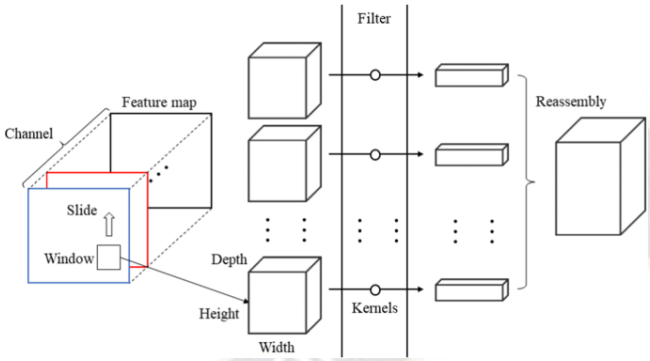

Figure4. Working of convolutional layer [72].

Parameter sharing is a key feature of the convolutional layer in a CNN, the feature detector's weights remain fixed as it moves across the image. During training, backpropagation and gradient descent are used to modify certain parameters, such as weight values. However, three hyperparameters need to be set before training the neural network begins, and they affect the volume size of the output. These hyperparameters are:

- **Depth:** The number of filters in the convolutional layer.
- **Stride:** The number of pixels to move the filter on each iteration.
- **Padding:** The number of pixels to add to each side of the input volume. The output volume size is calculated as follows:

$$\text{Output Volume Size} = ((W - F + 2P) / S) + 1 \quad (1)$$

Where W is the input volume size, F is the receptive field size of the filter, P is the amount of padding, and S is the stride. Setting these hyperparameters correctly is essential for the CNN to learn the features of the image effectively [76]. A more comprehensive study on the trends during the pandemic was done in [31] where they explored on whether India's vulnerability to cyberbullying was impacted by the COVID-19 outbreak. According to their statistics, stalking (71.21 %) was the most common kind of cyberbullying, followed by making disparaging remarks online (64.39%), releasing images or videos on the internet (41.67%), and harassing others (21.97%). This implies that a number of people have experienced more than one sort of cyberbullying. Although there has been an increase in cyberbullying incidents, The patterns for these statistics on cyberbullying and measures taken against perpetrators remain constant both prior to and during the COVID-19 epidemic. This implies that there is no effort being taken to address this issue and there is a need for an automatic detection system to protect social media users from being susceptible to such forms of harassment.

2) Pooling layer: Pooling layers are often referred to as downsampling, are used to reduce dimensionality and reduce the input's total number of parameters. Comparable to the convolutional layer, the pooling function runs a filter across the whole input with no weights [75] [77]. Instead, an aggregation function is used by a kernel when it is applied to a receptive field in order to aggregate the values and produce an output array. Max pooling and average pooling are the two most used pooling techniques. When the filter runs over the input in max pooling, the highest value in the receptive field is chosen and sent to the output array. The value in the receptive field that is closest to the average is chosen by average pooling, on the other hand, and it is sent to the output array. Although much information is lost in the pooling layer, it has several benefits to CNN, including reducing complexity, improving efficiency, and limiting the risk of overfitting [78].

3) Fully Connected (FC) Layer: As the name suggests, the FC layer connects every neuron in the preceding layer to every neuron in the output layer. In contrast, convolutional and pooling layers only connect some neurons to others. Based on the characteristics that were gathered using the various filters in the preceding levels, this layer classifies data. While ReLu functions are commonly used in convolutional and pooling layers, generally, FC layers use the SoftMax activation function to precisely identify inputs, resulting in a probability value ranging from 0 to 1. [76] [79].

B. Recurrent Neural Networks (RNNs)

A RNN is specifically designed to process data sequences with the time step index ranging from as seen in Figure 5. When working with sequential inputs, like speech and language, RNNs are usually the more effective solution. Knowing the terms before a given word in a phrase is crucial if you want to forecast it in an NLP challenge. The term RNN stands for Recurrent Neural Networks because they perform the same task for every element in a sequence, and the outcome is dependent on previous computations. [80] [81]. RNNs may also have a memory that stores details about previous calculations [82] [83] [84].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

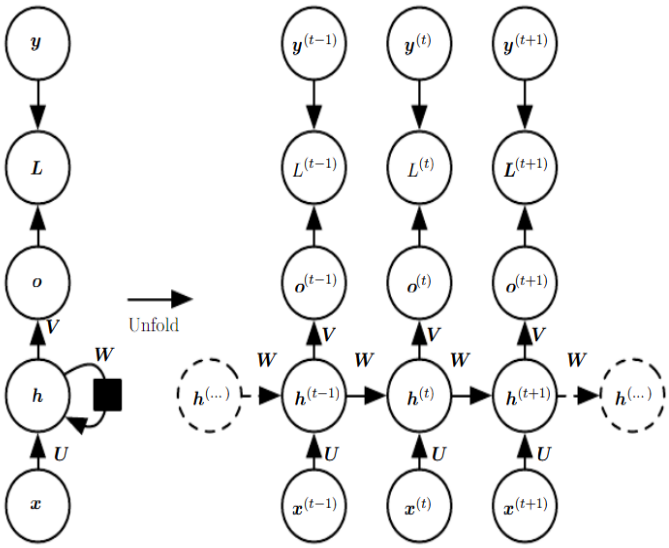

Figure5. Data Processing Increase.

- ***Hidden state h(t):*** functions as the network's "memory" and represents a hidden state at time t. The hidden state of the previous time step and the current input is used to determine the hidden state h(t): h(t) = f (U x(t) + W h(t-1)). It is assumed that the function f is a non-linear transformation like tanh or ReLU.
- ***Weights:*** The RNN model employs hidden-to-output connections, recurrent hidden-to-hidden connections, and input-to-hidden connections. These connections are established through weight matrices U, W, and V, respectively, which are shared across time t. This architecture of the RNN allows for the modelling of temporal dependencies in sequential data, making it an effective neural algorithm.

The Backpropagation Through Time (BPTT) technique is applied to recurrent neural networks to process sequence data, such as time series. In BPTT, the RNN processes the sequence one step at a time, and the gradients flow backward across time steps during the backpropagation process [85] [86]. This is done by unrolling the RNN over time, creating a feedforward neural network with multiple layers, as illustrated in Figure 6. The weights in this network are then updated using the standard Backpropagation algorithm. BPTT allows deep learning models to learn temporal dependencies in sequence data, making it useful in speech recognition, language modeling, and music generation [87]. BPTT unrolls each input timestep, computes error, adjusts weights, and calculates the cross-entropy loss 〖"E〗 _t " at each timestep t as explained in below equation 1. The overall loss is determined by adding up the losses at all time steps.

$$E_t = -y_t \log(\hat{y}_t) \quad (2)$$

In the above equationy_t denotes the actual output and yˆ_t represents the predicted output at timestamp t. The total loss with T-1 layers is represented as follows.

$$L = \sum_{j=0}^{T-1} L_j \quad (3)$$

When training an RNN, the significant task is to find the optimal weights that minimize the loss function. This involves adjusting the weights using an optimization algorithm such as gradient descent. By doing so, the RNN can learn to accurately predict the next element in a sequence and use contextual information from previous elements. Finding the optimal weights can be computationally intensive and requires much data. Still, it is essential for achieving high accuracy in RNN-based applications, but it results in vanishing or exploding gradients. Various versions of Recurrent Neural Networks (RNNs) have been developed, such as Bidirectional recurrent neural networks (BRNN) [88] [89], Long short-term memory (LSTM) [90], and Gated recurrent units (GRUs) [91] to address the exploding gradient issues [92].

V. DATASETS AND EVALUATION METRICS

A. Datasets

Deepfake detection is an essential area of research in today's digital age. To assist in the development of deepfake detection methods, several public benchmark datasets are available for experimental analysis. These datasets include UADFV, DFFD, FFHQ, 100K-Faces, CASIA-WebFace, VGGFace2, eye-blinking dataset, face swap-GAN, and Deepfake TIMIT. Each dataset has unique characteristics and features, making them suitable for different deepfake detection techniques. For example, FFHQ contains high-quality facial images, while DFFD includes real and fake facial images with different lighting and background conditions. Table 4 provides a comprehensive overview of these datasets, including the number of images, image resolution, and download links.

1) Flickr-Faces-HQ (FFHQ): The FFHQ dataset is a large group of high-quality images of human faces. It is a popular dataset for training and evaluating deep learning models for image generation, image editing, and face recognition. The FFHQ dataset contains 70,000 images, each of which is 1024x1024 pixels. The images are diverse regarding origin, race, and age. The dataset also includes A variety of accessories are available including caps, sunglasses, and eyeglasses. The FFHQ dataset was created by researchers at NVIDIA. The researchers collected the images from Flickr and then manually aligned and cropped them. They also removed any images that contained nudity or were not high quality. The FFHQ dataset is available for public download. It is a valuable resource for researchers and developers working to improve deep-learning models for image generation, editing, and face recognition.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

2) Fake Face Dataset (DFFD): The Diverse Fake Face Dataset (DFFD) is a massive dataset comprising real and fake faces. The dataset includes over 2.6 million images, with over 1.8 million fake faces generated through deepfake algorithms. The fake faces in the DFFD dataset contain different ages, genders, ethnicities, and image qualities, making the dataset incredibly diverse. The primary goal of creating the DFFD dataset was to eliminate the lack of diverse fake face datasets that can be used to train deepfake detection models. Previously available fake face datasets were often biased towards certain groups of people or specific types of deepfakes. However, the DFFD dataset was designed to be more diverse, aiming to train deepfake detection models that can effectively detect deepfakes in the real world.

3) CASIA-WebFace: The CASIA-WebFace dataset was introduced by Yi et al. [93] in their work on Learning Face Representation from Scratch. This dataset was created to facilitate face verification and face identification tasks. The CASIA-WebFace dataset comprises an enormous collection of 494,414 face images belonging to 10,575 real identities specifically collected from the web. The images were selected to represent various face variations, including facial expressions, poses, illumination conditions, and occlusions.

4) VGGFace2: Cao et al. [94] introduced a massive face recognition dataset designed to facilitate research in facial recognition. With an average of 362.6 photos per subject, the collection includes 3.31 million photos from 9131 subjects. These images were retrieved from Google Image Search and were specifically selected to capture a wide range of facial variations, including pose, age, illumination, ethnicity, and profession. The dataset includes images of individuals from diverse professions, such as actors, athletes, politicians, etc.

5) FaceForensic: The FaceForensics dataset [95] contains roughly 500,000 manipulated images from 1,004 videos for forensic benchmarking. It is divided into two subsets created using the Face2Face [3] reenactment approach. The dataset includes ground-truth masks and videos with a resolution of 854×480 or higher selected from YouTube and YouTube 8m datasets. The dataset is manually screened and comprises 1,408 training videos, 300 validation videos, and 300 testing videos.

6) Faceforensics++: FaceForensics++ [96] is a benchmark dataset for detecting realistic fake face images. It includes 1,000 carefully preferred videos from YouTube, with approximately 60% males and 40% females. Four different face alteration techniques have been applied to the videos after manual screening to ensure high quality. To provide a more reliable training dataset, the dataset offers a changed video along with a ground-truth mask that shows altered pixels for every input video.

7) UADFV: The University of Albany produced the useful synthetic dataset known as UADFV to aid in the detection of phoney face films using physiological indicators like eye blinking. The goal of the UADFV dataset is to address poor-quality synthesized videos lacking realistic features. This dataset consists of 49 authentic videos that were obtained from YouTube and 49 fraudulent videos that were created using the FakeApp mobile app 10, where the original faces of individuals were replaced with Nicolas Cage's face. Each video sequence is, on average, 11.14 seconds long and has a 294 × 500 pixels resolution. The UADFV dataset is vital for researchers and developers to improve face recognition technology and detect deepfake videos effectively. The UADFV dataset is a unique dataset that presents challenges in detecting deepfake videos that can be used to train machine learning or AI models.

8) Eye-Blinking Dataset: There are several eye-blinking datasets available for research and development purposes. Some of the most popular datasets include:

- Eye Blinking Dataset: This dataset contains 80 video clips of 20 individuals blinking, each lasting a few seconds. The videos were recorded under different lighting conditions and with different camera angles.
- Talking Face Dataset: This dataset contains over 5,000 video clips of people talking, each lasting 25 frames. The videos were recorded at a resolution of 720x576 pixels.
- mEBAL Dataset: This dataset contains over 340,000 eye blink images from 3,000 individuals. The images were recorded under different lighting conditions and with different head poses.
- UDF Dataset: This dataset contains 98 videos of real and fake faces generated using the FakeAPP deepfake algorithm. Each video is 11 seconds long on average and has a resolution of 294x500 pixels.

9) Deepfake TIMIT: The Deepfake TIMIT dataset is a collection of over 1,000 videos created by researchers at the Idiap Research Institute. The videos feature faces that have been swapped using the open-source GAN-based approach and are based on the TIMIT dataset, which contains audio-visual recordings of English speech. The Deepfake TIMIT dataset includes 16 different people, with each person having both real and fake videos. Two models are included in the dataset: a 64 × 64 model of lesser quality and a 128 × 128 input/output model of greater quality. Additionally, the author created 10 imaginary videos for each of the 32 subjects in each non-real video collection.

10) Faceswap-GAN: One of the publicly available datasets of deepfake videos created with Generative Adversarial Networks (GANs) is the Faceswap-GAN database, which was proposed by [97]. This dataset is composed of videos with varying levels of quality, ranging from low-quality to high-quality videos in 64 x 64 and 128 x 128-pixel dimensions, respectively. The dataset has 640 videos, approximately 200

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

frames each, and comprises manually chosen pairings of sixteen individuals from the VidTIMIT dataset [98]. The Faceswap-GAN database is a valuable contribution to deepfake detection and has been extensively utilized by scientists to create and assess deepfake detection models.

11) Celeb-DF: The Celeb-DF dataset, which was presented by Liu et al. [166], consists of 5,693 high-quality deepfake videos and 590 authentic videos. The dataset was created by collecting source videos from YouTube that feature subjects from different age groups, ethnicities, and genders. The dataset attempts to replicate various video quality real-world circumstances.

12) DFD: Google and Jigsaw collaborated to create the Deepfake Detection dataset (DDD/ DFD), a collection of deepfake videos and the advanced version of FaceForensics++ dataset. The dataset consists of 3,086 high-quality deepfake videos generated by employing 28 actors in 16 scenes. In addition to the deepfake videos, the dataset also includes 363 original videos. The project aimed to create a comprehensive dataset that can be used to train machine learning models to detect deepfakes accurately.

TABLE III. DEEPFAKE DETECTION PUBLICLY AVAILABLE DATASETS AND THEIR DESCRIPTION

S. No	Dataset Name	Original Data		Fake Data		Link
		Image	Video	Image	Video	
1	FFHQ	70000	-	-	-	https://github.com/NVlabs/ffhq-dataset
2	DFFD	58700	1000	240300	3000	https://cvlab.cse.msu.edu/dffd-dataset
3	CASIA-WebFace	494,414	-		-	https://github.com/happynear/AMSoftmax/issues/18
4	VGG Face2	3.31M	-	-	-	https://github.com/ox-vgg/vgg_face2
5	FaceForensics	500000	1004	521400	-	https://github.com/ondyari/FaceForensics
6	FaceForensics++	509900	1000	509900	4000	https://www.kaggle.com/sorokin/faceforensics/code
7	UADFV	17300	49	17300	49	https://github.com/LeeDongYeun/deepfake-detection
8	Eye Blinking	34000	5178	34000	5178	https://github.com/takhyun12/Dataset-of-Deepfakes
9	Deepfake TIMIT	-	-	-	620	https://www.idiap.ch/en/dataset/deepfaketimit
10	Celeb-DF	225400	590	2116800	5,693	https://github.com/yuezunli/celeb-deepfakeforensics
11	DFD	315400	363	2242700	3068	https://www.kaggle.com/c/deepfake-detection-challenge
12	DeeperForensic-1.0	-	-	-	60000	https://github.com/EndlessSora/DeeperForensics-1.0
13	SFD	156930	-	263123	150	https://github.com/yxlijun/S3FD.pytorch
14	WDF	11M	-	7314	707	https://github.com/deepfakeinthewild/deepfake-in-the-wild
15	Vision	344000	600	-	-	https://paperswithcode.com/dataset/vision
16	MANFA	8950	-	-	-	https://github.com/592McAvoy/fake-face-detection
17	TAMFA	7450	-	1500	-	https://www.sciencedirect.com/science/article/pii/S0957417419302350?via%3Dihub
18	FakeET	-	363	-	480	https://arxiv.org/abs/2006.06961
19	FFW	53000	150	-	-	https://github.com/AliKhoda/FFW
20	OF	16000	-	173000	-	https://paperswithcode.com/dataset/openforensics
21	FS	5000	1000	-	-	https://github.com/maum-ai/faceshifter

B. Evaluation Metrics

The evaluation metrics used to assess the performance of deepfake detection models are accuracy, f1 Score, recall, precision, Area Under the Curve (AUC), Receiver Operating Characteristic (ROC), False Acceptance Rate (FAR), False Rejection Rate (FRR), Error Rate (ER), Equal Error Rate (ERR), and Mean Absolute Error (MAE) as illustrated in table 5. The accuracy and AUC-ROC are usually primary metrics, as they provide an overall measure of the model's ability to detect deepfakes. These metrics serve as benchmark measures for evaluating the performance of different models. However, the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

other metrics, such as Recall, Precision, F1 Score, ERR, and FRR, provide more granular insights into the models' strengths and weaknesses. For example, The recall metric quantifies the percentage of genuine deepfakes that the algorithm accurately detects. Precision, on the other hand, measures the proportion of accurate deepfake detections. The F1 Score considers both Recall and Precision and provides a balance between the two. The ERR and FRR measures provide information on the misidentification rate of genuine videos as deepfakes, a crucial issue in deepfake detection. MAE is a statistical metric that measures the average magnitude of prediction errors, regardless of their direction. It's calculated by averaging the absolute differences between predicted and actual values and used to evaluate regression models.

TABLEV. EVALUATION METRICS USED FOR MEASURING THE PERFORMANCE OF THE DEEPFAKE DETECTION AND GENERATION MODELS.

S. No	Evaluation Metric	Formula
1	Accuracy	$Accuracy = \frac{TN + TP}{TN + FN + FP + FN}$
2	Precision	$Precision = \frac{TP}{TP + FP}$
3	Recall	$Recall = \frac{TP}{TP + FN}$
4	F1-score	$F1 - score = \frac{2(Precision \times Recall)}{Precision + Recall}$
5	FAR	$FAR = \frac{FP}{TP + FN}$
6	FRR	$FRR = \frac{FP}{FP + FN}$
7	ER	$ER = 1 - Accuracy$
8	ERR	$ERR = \frac{FAR + FRR}{2}$
9	MAE	$MAE = \frac{\lvert(actual - predicted)\rvert}{n}$

VI. APPLICATIONS OF DEEP FAKE

Deepfake technology has both positive and negative applications. The unethical use of Deepfake technology can have harmful consequences on our society, both in the short and long term. Those who frequently use social media are at a high risk of falling victim to Deepfakes. It's true that when used appropriately, Deepfake technology can bring about positive results. Below, I will describe in detail the negative and positive applications of Deepfake technology.

A. Negative Applications

In recent years, Deepfake technology and related technologies have expanded rapidly. Unfortunately, it has ample applications for malicious purposes, especially against celebrities and political leaders. Creating Deepfake content can have serious consequences, such as revenge, blackmail, and identity theft. There are thousands of videos on Deepfake, most of which are pornographic ones featuring women without their consent. Deepfake technology is most frequently used to create pornography of famous females, and it is especially popular with Hollywood actresses.

Additionally, a program that can instantly make a woman nude was developed in 2018 and is frequently used for malicious harassment of women. Deepfake technology has unfortunately been used for malicious purposes, such as creating fake recordings of politicians and world leaders. This has the potential to be a major threat to worldwide peace and stability. Many world leaders, such as former US President Barack Obama, current US President Donald Trump, US politician Nancy Pelosi, and German Chancellor Angela Merkel, have been targeted by fake videos. Even Facebook founder Mark Zuckerberg has faced similar incidents. Deepfake technology is also widely utilized for distinct areas, such as art, film organization, and social media.

Below is a list of some of the negative applications

1) Fraud: One of the major concerns with Deepfake technology is its potential for fraud. It has the ability to produce fictitious recordings of people saying or doing things they have never said or done in order to damage their reputation., steal their identity, or commit financial fraud.

2) Misinformation: Deepfake technology can also spread misinformation and disinformation, which can have severe consequences. It can influence elections, sow discord in society, or damage someone's reputation.

3) Harassment and bullying: Deepfakes have the potential to be used as a tool for harassment and bullying. For instance, an individual may create a deepfake video of another person doing or saying something embarrassing and then circulate it over the internet. This could lead to a severe impact on the victim's psychological well-being and social life.

4) Censorship: Deepfakes could be employed to censor or manipulate information. For instance, a government may use deepfakes to generate fabricated videos of its adversaries, confusing the public about what is authentic and what is not.

B. Positive Applications

While this technology is often employed maliciously, it can also have positive applications in various sectors. The ability to create Deepfakes has greatly increased, no longer being restricted to experts. The beneficial use of this technology now days are becoming increasingly prevalent. This technology is widely used to create new artwork, engage audiences, and offer unique experience. The Dalí Museum in St. Petersburg, Florida, recently provided its visitors with an opportunity to interact more intimately with the life and work of Salvador Dalí through artificial intelligence. Visitors could meet the great artist and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

learn more about his personality in an interactive experience. Deepfake technology is now being utilized for advertising and business purposes as well. Technologists use Deepfakes to replicate famous artwork, for instance, making videos of the well-known personalities Mona Lisa painting using the image [225]. The film industry can save a vast amount of money and time by utilizing the capabilities of Deepfake technology for video editing instead of reshooting. There are also many positive examples of this technology, the famous footballer David Beckham advocates for a malaria campaign and speaks in nine different languages. Additionally, Deepfakes can have positive impacts in the education sector as well.

Some of the positive applications of Deepfake technology include:

1) Art and entertainment: Deepfakes can be utilized to create innovative forms of art and entertainment, including music, movies, and video games. For instance, deepfakes have been employed to create realistic-looking music videos of deceased artists or to produce new dialogue for existing movies.

2) Education: Using deepfakes, educational materials may become more dynamic and captivating. For instance, deepfakes can be employed to create virtual simulations of historical events, allowing learners to experience the past in a more immersive way. Additionally, deepfakes can personalize learning experiences, making learning more accessible and effective for individuals with different learning styles and preferences.

3) Research: Deepfakes can be used to research human behaviour and psychology. For instance, deepfakes can be employed to study how people respond to different forms of persuasion or to comprehend the effect of fake news on public opinion.

4) Medical: Deepfakes can be used to enhance medical care. For instance, deepfakes can train surgeons on complex processes ensure that they have an appropriate environment in which to improve their abilities. Additionally, deepfakes can create personalized simulations of diseases, allowing doctors to understand and treat their patients better.

5) Accessibility: Deepfakes can make media more accessible to individuals with disabilities. For instance, deepfakes can generate video audio descriptions, making it possible for individuals with visual impairments to enjoy and understand the content. Additionally, deepfakes can be used to create transcripts of conversations, making it easier for individuals with hearing impairments to follow the dialogue.

VII. CHALLENGES AND FUTURE DIRECTIONS

The current findings and challenges related to Deepfakes are significant, as they provide valuable insights for creating more effective and difficult-to-detect Deepfakes in the future. With improved Deepfake generation methods, detecting Deepfakes accurately and efficiently will pose significant challenges. Innovative and rigorous efforts are required to develop more advanced and comprehensive detection mechanisms to address this. Such work is critical to mitigating the potential harms posed by Deepfakes and ensuring the integrity of digital media.

A. Deepfake Generation

Despite major attempts to improve Currently, there are several challenges that need to be addressed in order to improve the visual quality of Deepfakes. There are several issues to address, including lack of realism in facial features, inconsistent timing, limited lighting options, unrealistic hand movements, and potential for identity theft are some issues associated with Deepfake generation, as shown in figure 6.

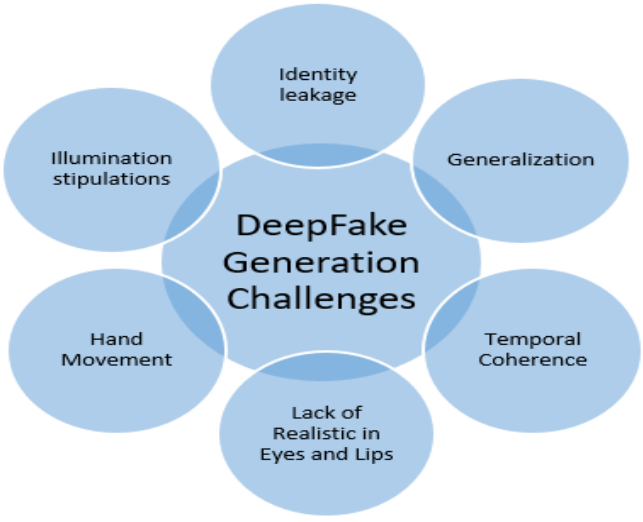

Figure*6. Deepfake generation challenges*

1) Generalization: Generative models' characteristics heavily depend on the dataset used during training. The output reflects the learned characteristics of that dataset, and its quality depends on the dataset's size. However, training a convincing model can be time-consuming, and finding enough data for a single victim can be challenging. Additionally, retraining the model for every distinct target identification requires a lot of resources.

2) Lack of high-resolution images: With the advent of Least Squares GAN (LS-GAN) [40], a technique for creating high-resolution images, there has been limited progress in advancing the synthesis of entirely fake images at even greater resolutions as shown in figure 7. This is significant because, as high-definition display resolutions continue to develop for devices such as phones and computers, high resolution may no longer suffice in the near future.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Figure8: Representation of a few examples of high-resolution images generated through LS-GAN [40]

3) Temporal coherence: Another limitation of DeepFake frameworks is the presence of anomalies that are obvious, including jittering and flashing between frames. These issues arise due to the frameworks' inability to consider temporal consistency while generating each frame. To address these limitations, researchers have proposed several approaches, such as incorporating contextual information into the generator or discriminator, employing temporal coherence losses as depicted in Figure 8, utilizing recurrent neural networks (RNNs), or combining multiple approaches to improve output quality.

4) Lack of diversified dataset: Fake datasets are focused on expanding the scale of data with limited consideration for diversity in video quality. The latest DeeperForensics-1.0 dataset has significantly progressed by the addition of disturbances like noise, contrast adjustments, JPEG compression, and Gaussian blur. The disturbances, on the other hand, are fabricated post-processing image-level degradations. We hope to see more organic degradations, such as bit-rate variations and codec choices, incorporated into upcoming datasets like DeeperForensics-2.0.

5) Illumination provision: Many of the accessible DeepFake datasets are created in an organized environment with a preset backdrop and lighting conditions. are kept the same throughout the process. However, when the lighting conditions suddenly shift in indoor and outdoor scenarios, it can cause There are differences in color and other abnormalities in the outcome.

6) Lack of hand movement generation: Another issue with Deepfake models is their inability to reflect emotions expressed through hand movements accurately. This is due to the limited availability of datasets that capture this type of expression, making it challenging to produce this type of Deepfake.

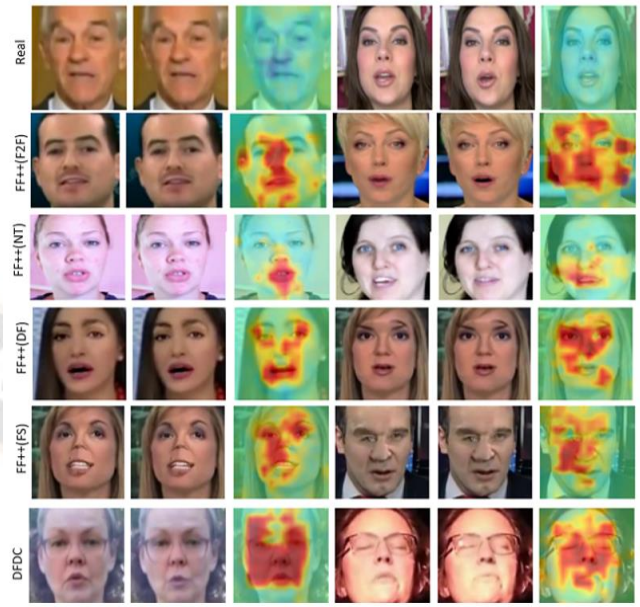

Figure 9: Representation of temporal coherence defects from various datasets [99]

7) Lack of realistic eyes and lip movement generation: The main challenges in producing eye and lip synchronization-based Deepfakes are the absence of genuine emotions, pauses, and changes in the target's speech rate. In addition, eye blinking abnormalities are another common issue that can arise in DeepFake-generated videos. The Figure 9 represents the original and fake generated video, and it is observed that the blinking of eyes is detected within 6 [100].

Figure10. Representation of detecting eye blinking from video [100]

8) Identity leakage: Maintaining target identity in face reenactment tasks can be a significant challenge when using a different source identity to drive the target expressions. This is because there is often a notable discrepancy between the target and driving identities. As a result, the facial data of the driving identity may only be partially transmitted to the manufactured face, leading to discrepancies in identity. This can occur when the training data is sourced from a single or many identities, yet the data pairing is performed on the same identity.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

B. Deepfake Detection

While there has been progress in improving the performance of Deepfake detectors, as demonstrated in figure 11, there are still challenges that need to be addressed. One of the primary challenges is the limited availability of datasets, which poses difficulties in developing accurate detection algorithms. Additionally, unknown types of attacks on media make it challenging to identify Deepfakes. In order to enhance the precision of detecting Deepfakes, it is imperative to address the challenges posed by temporal aggregation and unlabelled data. These issues are critical factors that hinder the accurate identification of Deepfakes. Therefore, it is crucial to develop mechanisms to overcome these challenges so that the detection of Deepfakes can be improved.

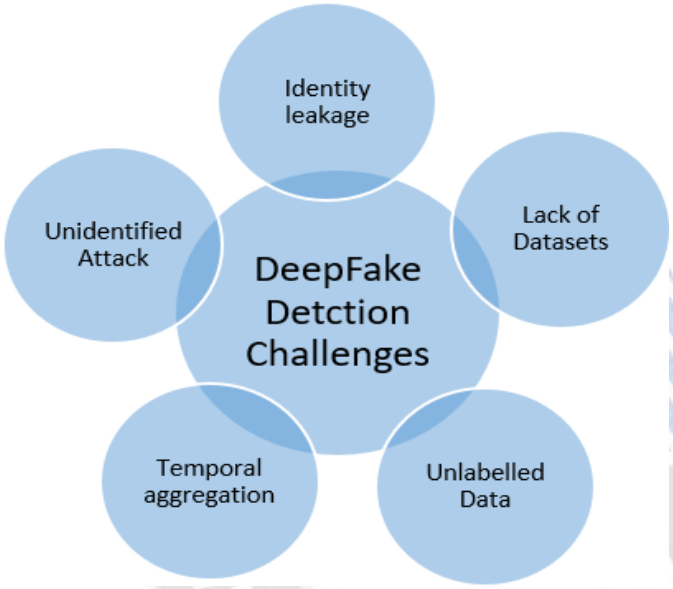

Figure11. *Deepfake detection challenges*

1) Lack of AI-synthesized datasets: Detecting Deepfakes is challenging due to the lack of consensus on image datasets used for evaluation and the need for synthetic fake image datasets. Prior studies have used GANs to create image datasets for testing methods in detecting Deepfakes in still images. However, the authenticity of these synthetic images is uncertain, and a standardized dataset is lacking. To advance research in this field, a public GAN-generated dataset of counterfeit images is essential, enabling consistent assessment of detection methods and improving accuracy and reliability of results.

2) Unlabelled data: Creating a reliable DeepFake detection model with a small and unlabelled dataset can be a daunting task, especially when the dataset requires the addition of scores that correspond to the type of forgery used. This challenge is amplified in the fields of journalism and law enforcement where data availability is usually limited. While most DeepFake detection models rely on large datasets for training and use deep learning approaches, they often lack transparency due to their black-box nature, making it difficult to understand how they operate. Therefore, more research is necessary to gain a deeper understanding of the complexities of designing DeepFake detection models for real-world applications.

3) Temporal Aggregation: Presently, DeepFake detection algorithms rely on binary frame-level classification to assess the authenticity of video frames. However, this approach has certain limitations as it does not consider the potential impact of interframe temporal consistency. This can result in temporal abnormalities and the presence of real and artificial frames in consecutive intervals. Additionally, the methods used in these algorithms require an additional step to calculate the video integrity score, which must be integrated for each frame to obtain the final outcome. These limitations highlight the need for more sophisticated algorithms to account for temporal information and produce more accurate and reliable results.

4) Generalization: Detecting Deepfakes that are not yet known to the system is a significant challenge against detecting fake videos. Many researchers are working on developing more generalized methods to tackle this challenge. However, most current studies have only been tested on basic datasets such as Face Forensics++. In order to enhance the efficacy and resilience of detection techniques, it is essential to prioritize the exploration of intricate and varied datasets in future research. This approach can potentially lead to the development of more robust and reliable methods of detection. Therefore, it is imperative that future research endeavors focus on the utilization of sophisticated and diverse datasets.

5) Capability and robustness: In order to develop a practical Deepfake detector that can be effectively utilized in real-world scenarios, it is imperative to enhance its ability to detect fake content across a broad range of situations. This involves fortifying its resistance to various Deepfake manipulations, including basic alterations and malicious attacks, while also providing an easily comprehensible explanation of its functionality. Recent research in Deepfake detection has indicated that only a handful of studies have thoroughly evaluated the efficacy of their approach from all three perspectives. Moving forward, comprehensive attention must be given to all three essential factors in the development of Deepfake detectors, in order to ensure optimal reliability and efficiency.

6) Unknown type of attack: Detecting Deepfakes is challenging when it comes to unknown types of attacks, such as Carlini WagnerL2 norm attack (CW-L2) and Fast Gradient Sign

Method (FGSM). These attacks trick classifiers into producing inaccurate outputs. Adversarial perturbations can make Deepfakes appear real, so it's important to develop robust detection models to withstand such attacks.

VIII. CONCLUSION

This survey states a comprehensive analysis of the research work on Deepfake, a complex and intricate topic that has garnered significant attention in recent years. A wide range of aspects related to Deepfake, including deepfake generation and detection, are presented. This review provides a detailed analysis of the different types of deepfakes and outlines the various methodologies used in detecting them, including traditional and deep learning-based approaches. Additionally, it provides a comprehensive overview of the resources available for generating deepfakes, such as tools, datasets, and evaluation metrics to measure the performance of deepfake-generated and detector models. It provides an in-depth understanding of the challenges and opportunities associated with deepfake technology and explores its potential applications and implications. Overall, this survey offers a valuable resource for researchers, academics, and practitioners interested in Deepfake and its related topics.

REFERENCES

[1] M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, "Deepfake detection: A systematic literature review," IEEE access 10, 2020.

[2] J. Hui, "How deep learning fakes videos (Deepfake) and how to detect it?," Medium Corporation 28, 2018.

[3] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt and M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets," Advances in neural information processing systems 27, 2014.

[5] S. Lyu, "Detecting deepfake videos in the blink of an eye," The Conversation 29, 2018.

[6] B. Marr, "The best (and scariest) examples of AI-enabled deepfakes," Forbes. https://cutt. ly/vK0OcsP, 2019.

[7] A. Roets, "Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions," Intelligence 65, 2017.

[8] M. Westerlund, "The emergence of deepfake technology: A review," Technology innovation management review 9, 2019.

[9] A. M. Almars, "Deepfakes detection techniques using deep learning: a survey," Journal of Computer and Communications 9, 2021.

[10] R. Rafique, R. Gantassi, R. Amin, J. Frnda, A. Mustapha and A. H. Alshehri, "Deep fake detection and classification using error-level analysis and deep learning," Scientific Reports 13, 2023.

[11] H. F. Shahzad, F. Rustam, E. S. Flores, J. L. V. Mazón, I. d. l. T. Diez and I. Ashraf, "A Review of Image Processing Techniques for Deepfakes," Sensors 22, 2022.

[12] J. Chai, H. Zeng, A. Li and E. W. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," Machine Learning with Applications 6, 2021.

[13] M. S. Rana and A. H. Sung, "Deepfakestack: A deep ensemble-based learning technique for deepfake detection," In 2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom), 2020.

[14] J. Liu, K. Zhu, W. Lu, X. Luo and X. Zhao, "A lightweight 3D convolutional neural network for deepfake detection," International Journal of Intelligent Systems 36, 2021.

[15] D. Wodajo and S. Atnafu, "Deepfake video detection using convolutional vision transformer," arXiv preprint arXiv:2102.11126, 2021.

[16] D. a. E. J. D. Güera, "Deepfake video detection using recurrent neural networks," In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), 2018.

[17] S. Agarwal and L. R. Varshney, "Limits of deepfake detection: A robust estimation viewpoint," arXiv preprint arXiv:1905.03493, 2019.

[18] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," arXiv preprint arXiv:1812.08685, 2018.

[19] Y. Li, M.-C. Chang and S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking," in In 2018 IEEE International workshop on information forensics and security (WIFS), 2018.

[20] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," arXiv preprint arXiv:1812.08685, 2018.

[21] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," arXiv preprint arXiv:1811.00656, 2018.

[22] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in In Proceedings of the European conference on computer vision (ECCV) , 2018.

[23] Y. Nirkin, Y. Keller and T. Hassner, "Fsgan: Subject agnostic face swapping and reenactment," in In Proceedings of the IEEE/CVF international conference on computer vision, 2018.

[24] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[25] J. Zhang, X. Zeng, M. Wang, Y. Pan, L. Liu, Y. Liu, Y. Ding and C. Fan, "Freenet: Multi-identity face reenactment," in In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.

[26] I. Korshunova, W. Shi, J. Dambre and L. Theis, "Fast face-swap using convolutional neural networks," in In Proceedings of the IEEE international conference on computer vision, 2017.

[27] G. B. Huang, M. Mattar, T. Berg and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition, 2008.

[28] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta and A. A. e. al., "Photo-realistic single image super-

resolution using a generative adversarial network," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.

[29] 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 In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

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

[31] R. Natsume, T. Yatagawa and S. Morishima, "Fsnet: An identity-aware generative model for image-based face swapping," in In Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, 2019.

[32] R. Natsume, T. Yatagawa and S. Morishima, "Rsgan: face swapping and editing using face and hair representation in latent spaces," arXiv preprint arXiv:1804.03447, 2018.

[33] Q. Sun, A. Tewari, W. Xu, M. Fritz, C. Theobalt and B. Schiele, "A hybrid model for identity obfuscation by face replacement," in In Proceedings of the European conference on computer vision (ECCV).

[34] C. Fu, Y. Hu, X. Wu, G. Wang, Q. Zhang and R. He, "High-fidelity face manipulation with extreme poses and expressions," in IEEE Transactions on Information Forensics and Security, 2021.

[35] Y. Wang, P. Bilinski, F. Bremond and A. Dantcheva, "Imaginator: Conditional spatio-temporal gan for video generation," in In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020.

[36] S. Tripathy, J. Kannala and E. Rahtu, "Icface: Interpretable and controllable face reenactment using gans," in In Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2020.

[37] 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.

[38] E. Zakharov, A. Shysheya, E. Burkov and V. Lempitsky, "Few-shot adversarial learning of realistic neural talking head models," in In Proceedings of the IEEE/CVF international conference on computer vision, 2019.

[39] T.-C. Wang, M.-Y. Liu, A. Tao, G. Liu, J. Kautz and B. Catanzaro, "Few-shot video-to-video synthesis," arXiv preprint arXiv:1910.12713, 2019.

[40] Z. Liu, H. Hu, Z. Wang, K. Wang, J. Bai and S. Lian, "Video synthesis of human upper body with realistic face," in In 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct).

[41] W. Wu, Y. Zhang, C. Li, C. Qian and C. C. Loy, "Reenactgan: Learning to reenact faces via boundary transfer," in In Proceedings of the European conference on computer vision (ECCV), 2018.

[42] Y. Song, J. Zhu, D. Li, X. Wang and H. Qi, "Talking face generation by conditional recurrent adversarial network," arXiv preprint arXiv:1804.04786 , 2018.

[43] L. Tran, X. Yin and X. Liu, "Representation learning by rotating your faces," IEEE transactions on pattern analysis and machine intelligence , vol. 41, no. 12, pp. 3007-3021, 2018.

[44] A. Bansal, S. Ma, D. Ramanan and Y. Sheikh, "Recycle-gan: Unsupervised video retargeting," in In Proceedings of the European conference on computer vision (ECCV), 2018.

[45] R. Kumar, J. Sotelo, K. Kumar, A. D. Brebisson and Y. Bengio, "Obamanet: Photo-realistic lip-sync from text," arXiv preprint arXiv:1801.01442 , 2017.

[46] S. Suwajanakorn, S. M. Seitz and I. Kemelmacher-Shlizerman, "Synthesizing obama: learning lip sync from audio," ACM Transactions on Graphics (ToG), vol. 36, no. 4, pp. 1-13, 2017.

[47] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in In Proceedings of the IEEE international conference on computer vision, 2017.

[48] Y. Yu, Z. Gong, P. Zhong and J. Shan, "Unsupervised representation learning with deep convolutional neural network for remote sensing images," in In Image and Graphics: 9th International Conference, Shanghai, China, 2017.

[49] M. Afifi, M. A. Brubaker and M. S. Brown, "Histogan: Controlling colors of gan-generated and real images via color histograms," in In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021.

[50] X. H. D. M. Q. Y. W. a. E. K. W. Nie, "Urca-gan: Upsample residual channel-wise attention generative adversarial network for image-to-image translation," Neurocomputing , vol. 443, pp. 75-84, 2021.

[51] J. Zeng, X. Ma and K. Zhou, "Photo-realistic face age progression/regression using a single generative adversarial network," Neurocomputing , vol. 366, pp. 295-304, 2019.

[52] Y. Jo and J. Park, "Sc-fegan: Face editing generative adversarial network with user's sketch and color," in In Proceedings of the IEEE/CVF international conference on computer vision, 2019.

[53] Z. W. Z. M. K. S. S. a. X. C. He, "Attgan: Facial attribute editing by only changing what you want," IEEE transactions on image processing , vol. 28, no. 11, pp. 5464-5478., 2019.

[54] M. Liu, Y. Ding, M. Xia, X. Liu, E. Ding, W. Zuo and S. Wen, "Stgan: A unified selective transfer network for arbitrary image attribute editing," in In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019.

[55] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu and L. Lin, "Beautygan: Instance-level facial makeup transfer with deep generative adversarial network," in In Proceedings of the 26th ACM international conference on Multimedia, 2018.

[56] T. Xiao, J. Hong and J. Ma, "Elegant: Exchanging latent encodings with gan for transferring multiple face attributes," in In Proceedings of the European conference on computer vision (ECCV), 2018.

[57] T. Karras, S. Laine and T. Aila, "A style-based generator architecture for generative adversarial networks," in In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019.

[58] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen and T. Aila, "Analyzing and improving the image quality of stylegan," in In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.

[59] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz and B. Catanzaro, "Video-to-video synthesis," arXiv preprint arXiv:1808.06601 , 2018.

[60] T. Karras, T. Aila, S. Laine and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," arXiv preprint arXiv:1710.10196 , 2017.

[61] J. a. Y. L. Guo, "Attributes guided facial image completion," Neurocomputing , vol. 392, pp. 60-69, 2020.

[62] Y. Chen, S. Xia, J. Zhao, M. Jian, Y. Zhou, Q. Niu, R. Yao and D. Zhu, "Person image synthesis through siamese generative adversarial network," Neurocomputing , vol. 417, pp. 490-500, 2020.

[63] D. B. L. Z. K. J. Z. J. Z. a. Z. X. Ma, "Two birds with one stone: Transforming and generating facial images with iterative GAN," Neurocomputing , vol. 396, pp. 278-290, 2020.

[64] K. M. S. J. L. D. L. B. C. a. D. C. Aberman, "Deep video ‐ based performance cloning," In Computer Graphics Forum, vol. 38, no. 2, pp. 219-233, 2019.

[65] L. W. X. M. Z. H. K. F. B. M. H. W. W. a. C. T. Liu, "Neural rendering and reenactment of human actor videos," ACM Transactions on Graphics (TOG) , vol. 38, no. 5, pp. 1-14, 2019.

[66] Y. Zhou, Z. Wang, C. Fang, T. Bui and T. Berg, "Dance dance generation: Motion transfer for internet videos," in In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.

[67] C. Chan, S. Ginosar, T. Zhou and A. A. Efros, "Everybody dance now," in In Proceedings of the IEEE/CVF international conference on computer vision, 2019.

[68] H. P. G. A. T. W. X. J. T. M. N. P. P. C. R. M. Z. a. C. T. Kim, "Deep video portraits," ACM transactions on graphics (TOG) , vol. 37, no. 4, pp. 1-14, 2018.

[69] S. Tulyakov, M.-Y. Liu, X. Yang and J. Kautz, "Mocogan: Decomposing motion and content for video generation," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[70] G. Balakrishnan, A. Zhao, A. V. Dalca, F. Durand and J. Guttag, ""Synthesizing images of humans in unseen poses," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[71] A. Siarohin, E. Sangineto, S. Lathuiliere and N. Sebe, "Deformable gans for pose-based human image generation," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[72] N. Neverova, R. A. Guler and I. Kokkinos, "Dense pose transfer," in In Proceedings of the European conference on computer vision (ECCV), 2018.

[73] O. Fried, A. Tewari, M. Zollhöfer, A. Finkelstein, E. Shechtman, D. B. Goldman, K. Genova, Z. Jin, C. Theobalt and M. Agrawala, "Text-based editing of talking-head video," ACM Transactions on Graphics (TOG) , vol. 38, no. 4, pp. 1-14, 2019.

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

[75] X. Li, Y. Lang, Y. Chen, X. Mao, Y. He, S. Wang, H. Xue and Q. Lu, "Sharp multiple instance learning for deepfake video detection," In Proceedings of the 28th ACM international conference on multimedia, 2020.

[76] L. Guarnera, O. Giudice and S. Battiato, "Fighting deepfake by exposing the convolutional traces on images," IEEE Access 8, 2020.

[77] M. Bonomi, C. Pasquini and G. Boato, "Dynamic texture analysis for detecting fake faces in video sequences," Journal of Visual Communication and Image Representation 79, 2021.

[78] J. Hernandez-Ortega, R. Tolosana, J. Fierrez and A. Morales, "Deepfakeson-phys: Deepfakes detection based on heart rate estimation," arXiv preprint arXiv:2010.00400, 2020.

[79] X. Dong, J. Bao, D. Chen, W. Zhang, N. Yu, D. Chen, F. Wen and B. Guo, "Identity-driven deepfake detection," arXiv preprint arXiv:2012.03930, 2020.

[80] A. Chintha, B. Thai, S. J. Sohrawardi, K. Bhatt, A. Hickerson, M. Wright and R. Ptucha, "Recurrent convolutional structures for audio spoof and video deepfake detection," IEEE Journal of Selected Topics in Signal Processing 14, 2020.

[81] S.-Y. Wang, O. Wang, R. Zhang, A. Owens and A. A. Efros, "CNN-generated images are surprisingly easy to spot... for now," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.

[82] Z. Guo, G. Yang, J. Chen and X. Sun, "Fake face detection via adaptive manipulation traces extraction network," Computer Vision and Image Understanding 204, 2021. A. Gandhi and S. Jain, "Adversarial perturbations fool deepfake detectors," In 2020 international joint conference on neural networks (IJCNN), 2020.

[83] W. Zhang, C. Zhao and Y. Li, "A novel counterfeit feature extraction technique for exposing face-swap images based on deep learning and error level analysis," Entropy 22, 2020.

[84] U. A. Ciftci, I. Demir and L. Yin, "How do the hearts of deep fakes beat? deep fake source detection via interpreting residuals with biological signals," In 2020 IEEE international joint conference on biometrics (IJCB), 2020.

[85] B. L. Welch, "The generalization of ‘STUDENT'S’problem when several different population varlances are involved," Biometrika 34, 1947.

[86] J. Benet, "Ipfs-content addressed, versioned, p2p file system," arXiv preprint arXiv:1407.3561, 2014.

[87] F. Matern, C. Riess and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), 2019.

[88] 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.

[89] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano and H. Li, "Protecting World Leaders Against Deep Fakes," In CVPR workshops, 2019.

[90] K. Songsri-in and S. Zafeiriou, "Complement face forensic detection and localization with faciallandmarks," arXiv preprint arXiv:1910.05455, 2015.

[91] H. H. Nguyen, J. Yamagishi and I. Echizen, "Capsule-forensics: Using capsule networks to detect forged images and videos," In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.

[92] H. H. Nguyen, J. Yamagishi and I. Echizen, "Use of a capsule network to detect fake images and videos," arXiv preprint arXiv:1910.12467, 2019.

[93] S. J. Sohrawardi, A. Chintha, B. Thai, S. Seng, A. Hickerson, R. Ptucha and M. Wright, "Poster: Towards robust open-world detection of deepfakes," In Proceedings of the 2019 ACM SIGSAC conference on computer and communications security, 2019.

[94] I. Amerini, L. Galteri, R. Caldelli and A. D. Bimbo, "Deepfake video detection through optical flow based cnn," In Proceedings of the IEEE/CVF international conference on computer vision workshops, 2019.

[95] D. Cozzolino, J. Thies, A. Rössler, C. Riess, M. Nießner and L. Verdoliva, "Forensictransfer: Weakly-supervised domain adaptation for forgery detection," arXiv preprint arXiv:1812.02510, 2018.

[96] H. H. Nguyen, F. Fang, J. Yamagishi and I. Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos," In 2019 IEEE 10th international conference on biometrics theory, applications and systems (BTAS), 2019.

[97] M. Du, S. Pentyala, Y. Li and X. Hu, "Towards generalizable deepfake detection with locality-aware autoencoder," In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020.

[98] M. Du, S. Pentyala, Y. Li and X. Hu, "Towards generalizable deepfake detection with locality-aware autoencoder," In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020.

[99] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera and D. Manocha, "Emotions don't lie: An audio-visual deepfake detection method using affective cues," In Proceedings of the 28th ACM international conference on multimedia, 2020.

[100] L. Guarnera, O. Giudice and S. Battiato, "Deepfake detection by analyzing convolutional traces," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020.

[101] S. Agarwal and L. R. Varshney, "Limits of deepfake detection: A robust estimation viewpoint," arXiv preprint arXiv:1905.03493, 2019.

[102] N.-T. Do, I.-S. Na and S.-H. Kim, "Forensics face detection from GANs using convolutional neural network," ISITC 2018 , 2018.

[103] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," arXiv preprint arXiv:1812.08685, 2018.

[104] H. M. Nguyen and R. Derakhshani, "Eyebrow recognition for identifying deepfake videos," In 2020 international conference of the biometrics special interest group (BIOSIG), 2020.

[105] R. Durall, M. Keuper, F.-J. Pfreundt and J. Keuper, "Unmasking deepfakes with simple features," arXiv preprint arXiv:1911.00686, 2019.

[106] A. Kumar, A. Bhavsar and R. Verma, "Detecting deepfakes with metric learning," In 2020 8th international workshop on biometrics and forensics (IWBF), 2020.

[107] K. Chugh, P. Gupta, A. Dhall and R. Subramanian, "Not made for each other-audio-visual dissonance-based deepfake detection and localization," In Proceedings of the 28th ACM international conference on multimedia, 2020.

[108] H. Qi, Q. Guo, F. Juefei-Xu, X. Xie, L. Ma, W. Feng, Y. Liu and J. Zhao, "Deeprhythm: Exposing deepfakes with attentional visual heartbeat rhythms," In Proceedings of the 28th ACM international conference on multimedia, 2020.

[109] P. Kawa and P. Syga, "A note on deepfake detection with low-resources," arXiv preprint arXiv:2006.05183, 2020.

[110] A. Khodabakhsh and C. Busch, "A generalizable deepfake detector based on neural conditional distribution modelling," In 2020 international conference of the biometrics special interest group (BIOSIG), 2020.

[111] I. Ganiyusufoglu, L. M. Ngô, N. Savov, S. Karaoglu and T. Gevers, "Spatio-temporal features for generalized detection of deepfake videos," arXiv preprint arXiv:2010.11844, 2020.

[112] A. Singh, A. S. Saimbhi, N. Singh and M. Mittal, "DeepFake video detection: a time-distributed approach," SN Computer Science 1, 2020.

[113] I. Kukanov, J. Karttunen, H. Sillanpää and V. Hautamäki, "Cost sensitive optimization of deepfake detector," In 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2020.

[114] 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.

[115] X. Zhu, H. Wang, H. Fei, Z. Lei and S. Z. Li, "Face forgery detection by 3d decomposition," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021.

[116] X. Wang, T. Yao, S. Ding and L. Ma, "Face manipulation detection via auxiliary supervision," In Neural Information Processing: 27th International Conference, ICONIP 2020, 2020.

[117] M. T. Jafar, M. Ababneh, M. Al-Zoube and A. Elhassan, "Forensics and analysis of deepfake videos," In 2020 11th international conference on information and communication systems (ICICS), 2020.

[118] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong and W. Xia, "Learning self-consistency for deepfake detection," In Proceedings of the IEEE/CVF international conference on computer vision, 2021.

[119] L. Bondi, E. D. Cannas, P. Bestagini and S. Tubaro, "Training strategies and data augmentations in cnn-based deepfake video detection," In 2020 IEEE international workshop on information forensics and security (WIFS), 2020.

[120] Z. Hongmeng, Z. Zhiqiang, S. Lei, M. Xiuqing and W. Yuehan, "A detection method for deepfake hard compressed videos based on super-resolution reconstruction using CNN," In Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence, 2020.

[121] Z. Hongmeng, Z. Zhiqiang, S. Lei, M. Xiuqing and W. Yuehan, "A detection method for deepfake hard compressed videos based on super-resolution reconstruction using CNN," In Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence, 2020.

[122] J. Han and T. Gevers, "MMD based discriminative learning for face forgery detection," In Proceedings of the Asian Conference on Computer Vision, 2020.

[123] H. Dang, F. Liu, J. Stehouwer, X. Liu and A. K. Jain, "On the detection of digital face manipulation," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern recognition, 2020.

[124] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini and S. Tubaro, "Video face manipulation detection through

ensemble of cnns," In 2020 25th international conference on pattern recognition (ICPR), 2021.

[125] M. S. Rana and A. H. Sung, "Deepfakestack: A deep ensemble-based learning technique for deepfake detection," In 2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom), 2020.

[126] S. Tariq, S. Lee and S. S. Woo, "A convolutional lstm based residual network for deepfake video detection," arXiv preprint arXiv:2009.07480, 2020.

[127] I. Masi, A. Killekar, R. M. Mascarenhas, S. P. Gurudatt and W. AbdAlmageed, "Two-branch recurrent network for isolating deepfakes in videos," In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16, 2020.

[128] L. Trinh, M. Tsang, S. Rambhatla and Y. Liu, "Interpretable and trustworthy deepfake detection via dynamic prototypes," In Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021.

[129] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen and B. Guo, "Face x-ray for more general face forgery detection," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.

[130] Z. Chen and H. Yang, "Attentive semantic exploring for manipulated face detection," In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.

[131] X. Ding, Z. Raziei, E. C. Larson, E. V. Olinick, P. Krueger and M. Hahsler, "Swapped face detection using deep learning and subjective assessment," EURASIP Journal on Information Security 2020, 2020.

[132] D. M. Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horváth and E. B. e. al., "Deepfakes detection with automatic face weighting," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020.

[133] Z. Liu, X. Qi, J. Jia and P. H. Torr, "Real or fake: An empirical study and improved model for fake face detection," 2019.

[134] R. Durall, M. Keuper and J. Keuper, "Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.

[135] M. A. S. Habeeba, A. Lijiya and A. M. Chacko, "Detection of deepfakes using visual artifacts and neural network classifier," In Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2020, 2021.

[136] K. Zhu, B. Wu and B. Wang, "Deepfake detection with clustering-based embedding regularization," In 2020 IEEE fifth international conference on data science in cyberspace (DSC), 2020.

[137] P. Charitidis, G. Kordopatis-Zilos, S. Papadopoulos and I. Kompatsiaris, "Investigating the impact of pre-processing and prediction aggregation on the deepfake detection task," arXiv preprint arXiv:2006.07084, 2020.

[138] X. Li, K. Yu, S. Ji, Y. Wang, C. Wu and H. Xue, "Fighting against deepfake: Patch&pair convolutional neural networks (PPCNN)," In Companion Proceedings of the Web Conference 2020, 2020.

[139] D. Cozzolino, A. Rössler, J. Thies, M. Nießner and L. Verdoliva, "Id-reveal: Identity-aware deepfake video detection," In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[140] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera and D. Manocha, "Emotions don't lie: An audio-visual deepfake detection method using affective cues," In Proceedings of the 28th ACM international conference on multimedia, 2020.

[141] Y. Nirkin, L. Wolf, Y. Keller and T. Hassner, "DeepFake detection based on discrepancies between faces and their context," IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 2021.

[142] C.-M. Yu, C.-T. Chang and Y.-W. Ti, "Detecting deepfake-forged contents with separable convolutional neural network and image segmentation," arXiv preprint arXiv:1912.12184, 2019.

[143] D. Feng, X. Lu and X. Lin, "Deep detection for face manipulation," In Neural Information Processing: 27th International Conference, ICONIP 2020, 2020.

[144] L. Chai, D. Bau, S.-N. Lim and P. Isola, "What makes fake images detectable? understanding properties that generalize," In Computer Vision–ECCV 2020: 16th European Conference, 2020.

[145] X. Chang, J. Wu, T. Yang and G. Feng, "Deepfake face image detection based on improved VGG convolutional neural network," In 2020 39th chinese control conference (CCC), 2020.

[146] A. Khodabakhsh, R. Ramachandra, K. Raja, P. Wasnik and C. Busch, "Fake face detection methods: Can they be generalized?," In 2018 international conference of the biometrics special interest group (BIOSIG), 2018.

[147] X. Yang, Y. Li and S. Lyu, "Exposing deep fakes using inconsistent head poses," In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.

[148] X. Zhang, S. Karaman and S.-F. Chang, "Detecting and simulating artifacts in gan fake images," In 2019 IEEE international workshop on information forensics and security (WIFS), 2019.

[149] S. Fernandes, S. Raj, E. Ortiz, I. Vintila, M. Salter, G. Urosevic and S. Jha, "Predicting heart rate variations of deepfake videos using neural ode," In Proceedings of the IEEE/CVF international conference on computer vision workshops, 2019.

[150] J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath and A. K. Roy-Chowdhury, "Hybrid lstm and encoder–decoder architecture for detection of image forgeries," IEEE Transactions on Image Processing 28, 2019.

[151] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," In Proceedings of the IEEE/CVF international conference on computer vision, 2019.

[152] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," arXiv preprint arXiv:1811.00656, 2018.

[153] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," Interfaces (GUI) 3, 2019.

[154] T. Fernando, C. Fookes, S. Denman and S. Sridharan, "Exploiting human social cognition for the detection of fake and fraudulent faces via memory networks," arXiv preprint arXiv:1911.07844, 2019.

[155]L. M. Dang, S. I. Hassan, S. Im and H. Moon, "Face image manipulation detection based on a convolutional neural network," Expert Systems with Applications 129, 2019.

[156]C.-C. Hsu, Y.-X. Zhuang and C.-Y. Lee, "Deep fake image detection based on pairwise learning," Applied Sciences 10, 2020.

[157]A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Nießner, "Faceforensics: A large-scale video dataset for forgery detection in human faces," arXiv preprint arXiv:1803.09179, 2018.

[158]P. Zhou, X. Han, V. I. Morariu and L. S. Davis, "Two-stream neural networks for tampered face detection," In 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW), 2017.

[159]P. Zhou, X. Han, V. I. Morariu and L. S. Davis, "Learning rich features for image manipulation detection," In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[160]D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "Mesonet: a compact facial video forgery detection network," In 2018 IEEE international workshop on information forensics and security (WIFS), 2018.

[161]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), 2018.

[162]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, 2022.

[163]Y. Yu, Z. Gong, P. Zhong and J. Shan, "Unsupervised representation learning with deep convolutional neural network for remote sensing images," In Image and Graphics: 9th International Conference, ICIG 2017, 2017.

[164]V. Thambawita, J. L. Isaksen, S. A. Hicks, J. Ghouse, G. Ahlberg, A. Linneberg and N. G. e. al, "DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine," Scientific reports 11, 2021.

[165]D. P. Kingma and M. Welling, "An introduction to variational autoencoders," Foundations and Trends® in Machine Learning 12, 2019.

[166]V. Blanz, K. Scherbaum, T. Vetter and H. P. Seidel, "Exchanging faces in images, in computer graphics forum," computer graphics forum, 2004.

[167]B. Wang, Y. Li, X. Wu, Y. Ma, Z. Song and M. Wu, "Face forgery detection based on the improved siamese network," Security and Communication Networks 2022, 2022.

[168]C. Vaccari and A. Chadwick, "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news," Social Media+ Society 6, 2020.

[169]B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," Calif. L. Rev, 2019.

[170]A. Siarohin, E. Sangineto, S. Lathuiliere and N. Sebe, "Deformable gans for pose-based human image generation," In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[171]A. Siarohin, E. Sangineto, S. Lathuiliere and N. Sebe, "Deformable gans for pose-based human image generation," In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[172]J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," In Proceedings of the IEEE international conference on computer vision, 2017.

[173]G.-Y. Hao, H.-X. Yu and W.-S. Zheng, "MIXGAN: learning concepts from different domains for mixture generation," arXiv preprint arXiv:1807.01659, 2018.

[174]L. Verdoliva, "Media forensics and deepfakes: an overview," IEEE Journal of Selected Topics in Signal Processing 14, 2020.

[175]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), 2019.

[176]A. v. d. Oord, Y. Li and O. Vinyals, "Representation learning with contrastive predictive coding," arXiv preprint arXiv:1807.03748, 2018.

[177]D. P. Kingma and M. Welling, "An introduction to variational autoencoders," Foundations and Trends® in Machine Learning 12, 2019.

[178]A. v. d. Oord and N. Kalchbrenner, "Pixel rnn," 2016.

[179]A. V. d. Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals and A. Graves, "Conditional image generation with pixelcnn decoders," Advances in neural information processing systems 29, 2016.

[180]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, 2022.

[181]Y. Chen, Y. Zhao, W. Jia, L. Cao and X. Liu, "Adversarial-learning-based image-to-image transformation: A survey," Neurocomputing 411, 2020.

[182]L. Guarnera, O. Giudice and S. Battiato, "Deepfake detection by analyzing convolutional traces," in In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020.

[183]M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, "Deepfake detection: A systematic literature review," IEEE access , vol. 10, pp. 25494-25513, 2020.

[184]A. S. Abdulreda and A. J. Obaid, "A landscape view of deepfake techniques and detection methods," International Journal of Nonlinear Analysis and Applications, vol. 13, no. 1, pp. 745-755, 2022.

[185]D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "Mesonet: a compact facial video forgery detection network," in In 2018 IEEE international workshop on information forensics and security (WIFS), 2018.

[186]R. Tolosana, S. Romero-Tapiador, J. Fierrez and R. Vera-Rodriguez, "Deepfakes evolution: Analysis of facial regions and fake detection performance," in In international conference on pattern recognition, 2021.

[187]P. Zhou, X. Han, V. I. Morariu and L. S. Davis, "Two-stream neural networks for tampered face detection," in In 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW), 2017.

[188]L. Bondi, E. D. Cannas, P. Bestagini and S. Tubaro, "Training strategies and data augmentations in cnn-based deepfake video detection," in 2020 IEEE international workshop on information forensics and security (WIFS), 2020.

[189]J. Liu, K. Zhu, W. Lu, X. Luo and X. Zhao, "A lightweight 3D convolutional neural network for deepfake detection,"

International Journal of Intelligent Systems, vol. 36, no. 9, pp. 4990-5004., 2021.

[190] I. Amerini, L. Galteri, R. Caldelli and A. D. Bimbo, "Deepfake video detection through optical flow based cnn," in In Proceedings of the IEEE/CVF international conference on computer vision workshops, 2019.

[191] S. R. Ahmed, E. Sonuç, M. R. Ahmed and A. D. Duru, "Analysis survey on deepfake detection and recognition with convolutional neural networks," in In 2022 International Congress on Human-Computer Interaction, Robotic Applications (HORA), 2022.

[192] A. Deshmukh and S. B. Wankhade, "Deepfake detection approaches using deep learning: a systematic review," in Intelligent Computing and Networking: Proceedings of IC-ICN, 2020.

[193] S. A. Aduwala, M. Arigala, S. Desai, H. J. Quan and M. Eirinaki, "Deepfake Detection using GAN Discriminators," in In 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), 2021.

[194] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), 2018.

[195] D. M. H. H. S. K. Y. S. B. R. J. H. S. Montserrat and E. B. e. al, "Deepfakes detection with automatic face weighting," in In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020.

[196] L. Chua, T. Roska, L. O. Chua and T. Roska, "The CNN paradigm," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 40, no. 3, pp. 147-156, 1993.

[197] L. O. Chua, "CNN: A paradigm for complexity," World Scientific, vol. 31, 1998.

[198] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE transactions on neural networks and learning systems , 2021.

[199] S. Cong and Y. Zhou, "A review of convolutional neural network architectures and their optimizations," Artificial Intelligence Review , vol. 56, no. 3, pp. 1905-1969, 2023.

[200] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai and T. L. e. al, "Recent advances in convolutional neural networks," Pattern recognition , vol. 77, pp. 354-377, 2018.

[201] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," Neural computation, vol. 29, no. 9, pp. 2352-2449, 2017.

[202] Q. Zhao and L. D. Griffin, "Suppressing the unusual: towards robust cnns using symmetric activation functions," arXiv preprint arXiv:1603.05145 , 2016.

[203] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," in In 2017 international conference on engineering and technology (ICET), 2017.

[204] M. Z. S. X. J. J. P. a. Y. P. Sun, "Learning pooling for convolutional neural network," Neurocomputing , vol. 224, pp. 96-104, 2017.

[205] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," arXiv preprint arXiv:2009.07485, 2020.

[206] S. S. S. R. D. V. P. a. S. M. Basha, "Impact of fully connected layers on performance of convolutional neural networks for image classification," Neurocomputing , vol. 378, pp. 112-119, 2020.

[207] W. Zaremba, I. Sutskever and O. Vinyals, "Recurrent neural network regularization," arXiv preprint arXiv:1409.2329 , 2014.

[208] I. Goodfellow, Y. Bengio and A. Courville, Sequence modeling: recurrent and recursive nets, 2016, pp. 367-415.

[209] 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.

[210] K. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

[211] L. R. a. L. C. J. Medsker, "Recurrent neural networks," Design and Applications , vol. 5, p. 2, 2001.

[212] M. Boden, "A guide to recurrent neural networks and backpropagation," the Dallas project, vol. 2, no. 2, pp. 1-10, 2002.

[213] F. J. Pineda, "Generalization of back-propagation to recurrent neural networks," Physical review letters , vol. 59, no. 19, p. 2229, 1987.

[214] T. P. A. S. L. M. C. J. A. a. G. H. Lillicrap, "Backpropagation and the brain," Nature Reviews Neuroscience , vol. 21, no. 6, pp. 335-346, 2020.

[215] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE transactions on Signal Processing, vol. 45, no. 11, pp. 2673-2681, 1997.

[216] Z. Cheng, R. Lu, Z. Wang, H. Zhang, B. Chen, Z. Meng and X. Yuan, "BIRNAT: Bidirectional recurrent neural networks with adversarial training for video snapshot compressive imaging," in In European Conference on Computer Vision, 2020.

[217] Y. Yu, X. Si, C. Hu and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," Neural computation , vol. 31, no. 7, pp. 71235-1270, 2019.

[218] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in Gate-variants of gated recurrent unit (GRUIn 2017 IEEE 60th International Midwest Symposium on circuits and systems (MWSCAS), 2017.

[219] J. Hanson, Y. Yang, K. Paliwal and Y. Zhou, "Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks," Bioinformatics , vol. 33, no. 5, pp. 685-692, 2017.

[220] D. Yi, Z. Lei, S. Liao and S. Z. Li., "Learning face representation from scratch," arXiv preprint arXiv:1411.7923 , 2014.

[221] Q. Cao, L. Shen, W. Xie, O. M. Parkhi and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in In 2018 13th IEEE international conference on automatic face & gesture recognition , 2018.

[222] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Nießner, "Faceforensics: A large-scale video dataset for forgery detection in human faces," arXiv preprint arXiv:1803.09179, 2018.

[223] P. Korshunov and S. Marcel, "Vulnerability assessment and detection of deepfake videos," in In 2019 International Conference on Biometrics (ICB), 2019.

[224] C. Sanderson, "The vidtimit database," No. REP_WORK. IDIAP, 2002.

[225] M. Westerlund, "The emergence of deepfake technology: A review," Technology innovation management review 9, 2011.

[226] Y. Zheng, J. Bao, D. Chen, M. Zeng and F. Wen, "Exploring temporal coherence for more general video face forgery detection," in In Proceedings of the IEEE/CVF international conference on computer vision, 2021.

[227] A. Khalid, "Deepfake Videos Are a Far, Far Bigger Problem for Women," Quartz, 2020.

[228] E. J. Dickson, "Deepfake porn is still a threat, particularly for k-pop stars," Rolling Stone, 2019.