Journal of Computer and Communications, 2021, 9, 20-35
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Deepfakes Detection Techniques Using Deep Learning: A Survey

Abdulqader M. Almars

College of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia
Email: amars@taibahu.edu.sa

How to cite this paper: Almars, A.M. (2021) Deepfakes Detection Techniques Using Deep Learning: A Survey. *Journal of Computer and Communications*, **9**, 20-35.
https://doi.org/10.4236/jcc.2021.95003

Received: April 12, 2021
Accepted: May 16, 2021
Published: May 19, 2021

Abstract

Deep learning is an effective and useful technique that has been widely applied in a variety of fields, including computer vision, machine vision, and natural language processing. Deepfakes uses deep learning technology to manipulate images and videos of a person that humans cannot differentiate them from the real one. In recent years, many studies have been conducted to understand how deepfakes work and many approaches based on deep learning have been introduced to detect deepfakes videos or images. In this paper, we conduct a comprehensive review of deepfakes creation and detection technologies using deep learning approaches. In addition, we give a thorough analysis of various technologies and their application in deepfakes detection. Our study will be beneficial for researchers in this field as it will cover the recent state-of-art methods that discover deepfakes videos or images in social contents. In addition, it will help comparison with the existing works because of the detailed description of the latest methods and dataset used in this domain.

Keywords

Deepfakes, Deep Learning, Fake Detection, Social Media, Machine Learning

1. Introduction

With the technology becoming accessible to any user, lots of deepfake videos have been spread through social media. Deepfake refers to manipulated digital media such as images of videos where the image or video of a person is replaced with another person's likeness. In fact, deepfake is one of the increasingly serious issues in modern society. Deepfake has been frequently used to swipe faces of popular Hollywood celebrities over porn images videos deepfake was also used to produce misleading information and rumors for politicians [1] [2] [3]. In

2018, a fake video for Barack Obama was created to putting words he never uttered [4]. In addition, In the US 2020 election, deepfakes have already been used to manipulate Joe Biden videos showing his tongue out. These harmful uses of deepfakes can have a serious impact on our society and can also result in spreading miss leading information, especially on social media.

Generative adversarial networks (GANs) [5] are generative and sophisticated deep learning technologies that can be applied to generate fake images and videos that hard for a human to identify from the true ones. Those models are used to train on a data set and then create fake images and videos. This kind of deepfake model requires a large set of training data for those deepfake media. The larger data set, the more believable and realistic images and videos can be created by the model. In fact, the large availability of presidents and Hollywood celebrity's videos on social media can help individuals to produce realistic fake news and rumors that can bring a serious impact on our society.

Recent studies show that deepfake video and images have become heavily circulated through social channels. Detection of deepfake videos and images, therefore, has become increasingly critical and important. To encourage researchers, many organizations such as United States Defense Advanced Research Projects Agency (DARPA), Facebook Inc and Google launched a research initiative in attempting the detection and prevention of deepfake [6] [7]. As a result, many deep learning approaches such as long short-term memory (LSTM), recurrent neural network (RNN) and even the hybrid approaches has been proposed to in order to detect deepfakes images and videos and to bring up more research over this field [8] [9] [10] [11]. The current studies show that deep neural networks made a remarkable result in terms of detecting fake news and rumors in social media posts.

This work primarily focuses on providing a comprehensive study for deepfake detection using deep-learning methods such as Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Long short-term memory (LSTM). This survey will be useful and beneficial for researchers in this field as it will give: 1) details summary of the current research studies; 2) datasets used in this field; 3) the limitations of the current approaches and insights of future work. The contributions of our survey are summarized as follows.

- This is the first review that covers the current deep learning methods for deepfake detection and discusses their limitations.
- In this survey, we also cover all the challenges that are faced by the researchers in this area and provide an outlook of future directions.
- In this review, we summarize and present available annotated datasets that are used in this field.

The rest of the article is structured as follows: Section 2 summaries the related work, Section 3 contains deepfake Creation and deep learning Detection Techniques, Section 4 contains the public available dataset used in the Deepfake field, the challenges and the open issues are discussed in Section 5, Section 6 concludes the research.

2. Related Work

2.1. Basics of Artificial Neural Networks (ANNs)

The basic concept of Artificial Neural Networks (ANNs) is partially inspired by how the human brain functions. Figure 1 shows artificial neural networks architecture. Neural networks are multi layers networks that consist of a single input layer, one or multi hidden layers and one output layers. The input to neural networks is a set of input values [12]. The goal of neural networks is to predict and classify those values into predefined categories.

The first layer in neural network is the input layers which takes input values and pass them to the next layer [13]. In our example, the input values are x_1, x_2, x_3 and x_4. The second layer is the Hidden layers which a set of connected unites called artificial neurons (nodes). The edges that connect the neurons represents how all the neurons are interconnected and how can receive and send signals through multi layers. Each connection has a weight associated with it which represents the connections between two units. In our network, the 1st hidden layer consists of 3 neurons and the 2nd layer contains 4 neurons. Each neuron receives number of inputs from previous layer and a bias value. A bias value is an extra value which equal to 1. If a neuron has n inputs, it should have n weight values which can be represented by the following learning formula (Equations (1) and (2)):

$$z = x_1 w_1 + x_2 w_2 + x_3 w_3 + \cdots + x_n w_n + b * 1 \tag{1}$$

$$z = \sum_{n=1} x_n w_n + b \tag{2}$$

The third layer is the output layer which reads the output from previous layer and predicts the output values y_1, y_2 and y_3. The goal of the learning and predicting process is to adjust the connection weights between those units to reduce the error and predict the output values. Activation functions are used in neural network to determine the output values $z = \sum_{n=1} x_n w_n + b$ of the model. Activation function aims to normalize the values into a smaller range. Equations (3), (4) and (5) are the most common activation functions used in neural network.

$$\text{sigmoid}(z) = \frac{1}{1 + \exp(-z)} \tag{3}$$

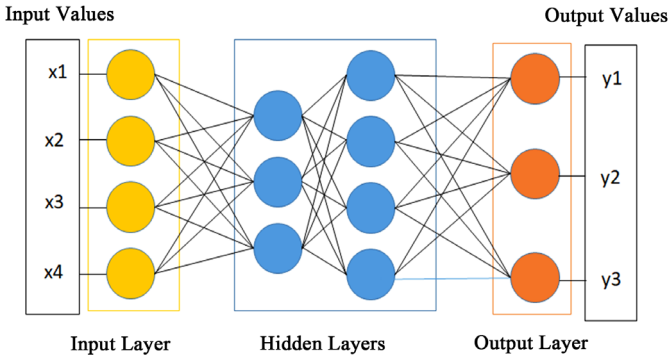

Figure 1. Artificial neural networks architecture.

$$\tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} - e^{-z}} \quad (4)$$

$$\mathrm{ReLU}(z) = \mathrm{Max}(0, z) \quad (5)$$

Sigmoid function Equation (3) is a widely used function that squashes values between a range [0, 1]. The tanh function Equation (4) is zero-centered which means it outputs values between [−1, 1] instead of [0, 1]. The rectified linear function (ReLU) Equation (5) is non-linear function that outputs the values directly if positives, otherwise, it will outputs zero. Compared with the other function's methods, ReLU has become the defuel activation function for many applications of neural networks as it easy to compute and fast to train.

To predict the input values of neural networks, the input values should be faded in a forward direction. This process of feeding inputs in this way is called Forward propagation. Thus, each hidden layer in neural network reads the inputs from previous layer, processes it thought the activation function and finally predicts the output values. Back propagation is a back forward process which aims at optimizing the weights to make sure that the neural network can correctly predict the outputs. To achieve this, stochastic gradient descent is used to reduce the error cost.

2.2. Deep Learning

Deep learning is a machine learning method based on the same idea of neural network [13] [14]. In deep learning, the word deep indicates the use of multiple hidden layers in the network. Inspired by artificial networks, the deep learning architecture uses an unbounded number of hidden layers of bounded size to extract higher information from raw input data. The number of hidden layers is determined based on the complexity of the training data [6]. More complex data requires more hidden layers to effectively produce the correct results. In recent years, deep learning has been used successfully in a variety of areas, including computer vision, audio processing, automatic translation, and natural language processing. Applying deep learning in these fields provides state-of-art results compared with the machine learning approaches. Deep learning also has shown promising results in deepfake detection. In literature, several techniques based on deep learning have been proposed including: 1) convolutional neural network (CNN); 2) recurrent neural network (RNN); 3) long short-term memory (LSTM). In the following sections, we briefly describe these techniques and then explain its implementation on deepfake discovery.

2.2.1. Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is the most commonly used deep neural network model. CNN, like neural networks, has an input and output layer, as well as one or more hidden layers. In CNN [15], the hidden layers first read the inputs from the first layer and then apply a convolution mathematical operation on the input values. Here, convolution indicates a matrix multiplication or other

dot product. After applying matrix multiplication, CNN uses the nonlinearity activation function such as Rectified Linear Unit (RELU) followed by additional convolutions such as pooling layers. The main goal of pooling layers is to reduce the dimensionality of the data by computing the outputs utilizing functions such as maximum pooling or average pooling.

2.2.2. Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) is another application of artificial neural network which is capable to learn features from sequence data [16]. Similar to neural networks, RNN is made up of several invisible layers, each of which has a weight and a bias. In RNN, the relations between nodes in a direct cycle graph that run in sequential order. One advantage of RRN is that it allows discovering temporal dynamic behavior. Compared with feed forward networks (FFN), RNN use an internal memory to store the sequences information from previous inputs, which makes it useful in a variety of areas, including natural language analysis and speech recognition. RNN can handle a temporal sequence by introducing a recurrent hidden state which captures dependencies of different time scales [17].

2.2.3. Long Short-Term Memory (LSTM)

LSTM [18] [19] is a type of artificial recurrent neural network (RNN) that handles long-term dependencies. LSMT contains feedback connections to learn the entire sequence of data. LSTM has been applied to many fields that based on time series data such as classifying, processing and making predictions. The common architecture of LSTM consists of: 1) input gate; 2) forget gate; 3) and an output gate. The cell state is long-term memory that remembers values from previous intervals and stores them in the LSTM cell. First, the input gate is responsible of selecting the values that should enter the cell state. The forget gate is reasonable of determining which information is to forget by applying a sigmoid function, which has a range of [0, 1]. The output gate determines which information in the current time should be considered in the next step.

3. Deepfake Generation and Detection

Deepfake is a technique that uses the Generative adversarial networks (GANs) methods to generate fictitious photographs and videos. In this section, we fist give an overview of the current applications and tools to create deepfake image and videos. Then, we discuss some deep learning detection techniques to overcome this issue.

3.1. Deepfake Generation

Generative adversarial networks (GANs) are a form of deep neural network that has been commonly used to generate deep fake. One advantage of GNAs is that it capable to learn from a set of training data set and create a sample of data with the same features and characteristics. For example, GANs can be used to swipe a

"real" image or the video of a person with that of a "fake" one [1]. The architecture of GANs consists of two neural networks components: an encoder and decoder. First, the model uses the encoder to train on a large data set to create fake data. Then, the decoder is used to learn the fake data from realistic data. However, this model requires a large amount data (images and videos) to generate realistic-looking faces. **Figure 2** shows the GNA architecture. As illustrated in the figure, the encoder first receives random inputs seeds to generate a fake sample. Those fake samples are used to train the decoder. The decoder is simply a binary classifier, and it takes the real samples and fake samples as inputs and then, decoder applies a SoftMax function to distinguish the realistic data from the fake one.

Many deepfake applications have already been around for quite a few years. FakeApp is the first method that has been used widely for deepfake creation. This FakeApp capable of swapping faces on videos using autoencoder-decoder pairing structure developed by a Reddit user [20] [21]. Similar to GANs, FakeApp consists of the autoencoder which is used to construct latent features of the human face images and, the decoder which is used to re-extract the features for the human face images. This simple technique is powerful as it capable to produce extremely realistic fake videos that hard for people to differentiate from the real one. VGGFace is another is another popular deepfake technique based on the generative adversarial network (GAN). The architecture of VGGFace [22] is improved by adding two layers called adversarial loss and perceptual lost. Those layers is added to autoencoder-decoder capture latent features of face images such as eye movements in order to produce more believable and realistic fake images.

CycleGAN [23] is a deepfake technique that extracts the characteristics of one image and produces another image with the same characteristics via the GAN architecture. This method applies cycle loss function that enables them to learn the latent features. Dissimilar from FakeApp, CycleGAN is unsupervised method that can perform image-to-image conversion without using paired examples. On other words, the model learns the features of a collection of images from the source and target that do not need to be related to each other's.

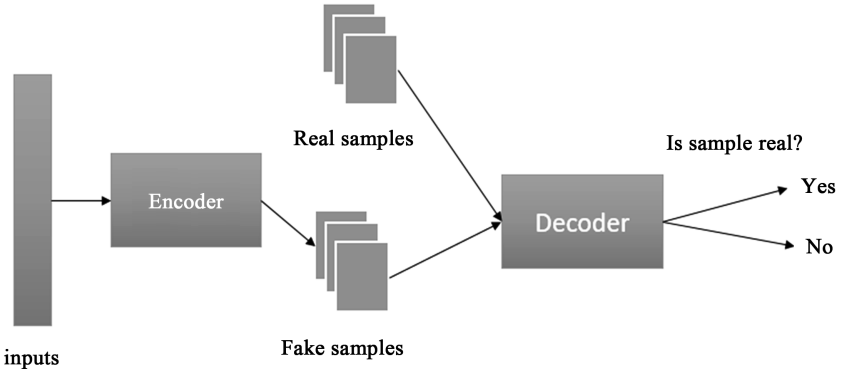

Figure 2. GNA architecture.

3.2. Deepfake Detection

Deep learning has achieved great success in deepfake detection. In this subsection below, we first discuss the Image Detection models using deep learning technologies and then Video Detection models are presented.

3.2.1. Image Detection Models

Different methods have been proposed to detect the GAN generated images using deep networks. Tariq *et al.* [24] suggested neural network-based methods for detecting fake GAN videos. This method employs pre-processing techniques to analyses the statistical features of image and enhances the detection of fake face image created by humans [25]. Nhu *et al.* [26] also introduces another approach based on a deep convolutional neural network for detecting fake image generated by GANs. The model firsts use a deep learning network to extract face features based on face recognition networks. Then, a fine-tuning step is used to make face features suitable for real/fake image detection. These methods produce good results from the contest validation data.

However, the majority of previous research ignores the critical issue of the forensics model's generalization capabilities. On other words, they use the same type of dataset to train and test their models. To tackle this problem, Xuan *et al.* [27] introduces a forensics convolutional neural network (CNN) that applies two image preprocessing steps to detect fake human images: Gaussian Blur and Gaussian Noise. The idea behind this model is to use preprocessing steps to neglect low level high frequency clues artifact in GAN images and improve high frequency pixel noise in low level pixel statistics. This enables the forensic classifier to learn more meaningful characteristics of real and false images, allowing it to better distinguish between real and fake image faces. The findings of the experiment reveal that the model can detect false images.

In addition to the traditional deepfake detection models, a hybrid approach was introduced to effectively detect the fake images [28] [29] [30]. Zhou *et al.* [29] for example proposed a two-stream network for detecting face tampering (see **Figure 3**). The face classification stream is used on GoogleNet [31] to train the model on tampered and authentic images. Then, the patch triplet stream is used to analysis features using steganalysis feature extractor and captures low

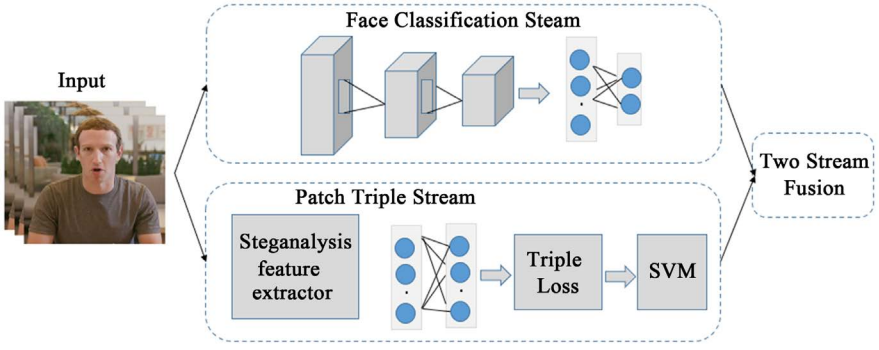

Figure 3. Two-stream neural networks.

level camera characteristics and local noise residuals. The experimental results show that this approach can learn both fake and real images. Another hybrid approach was introduced which use a pairwise-learning for deepfake image detection [30]. The approach first uses GANs to create and generate a fake image. Then, on the popular fake feature network (CFFN) generated by GANs, a pairwise-learning model is used to capture the discriminant information between the fake image and the real image. The evaluation results show that this approach can overcome the shortcomings of the existing state-of-the-art fake image detectors.

3.2.2. Video Detection Models

For the last years, deep learning methods have been successfully applied for fake image detection. However, the current deep learning methods for image cannot be directly applied for fake videos detection due to the availability of significant loss of frame information after video compression [32] [33]. In the subsection below, we have divided the related work in deepfake video detection into two main categories: biological singles analysis and spatial and temporal features analysis.

1) Biological Singles Analysis

Yuezun Li [34] presented a new approach based on natural network to detect Fake Face Videos. Compared with the previous work, this method considers eye blinking to detect fake videos, which is an important physical feature that can be used to distinguish the fake videos. To achieve that, this method uses a convolutional neural network (CNN) with a recursive neural network (RNN) to discover the physiological signals such eye movement and blinking. Then, the model uses a binary classifier to detect the close and open eyes state. This approach is tested with a dataset called eye-blinking crawled from the internet. The eye-blinking datasets is the first available dataset which specially designed for the eye-blinking detection. The experiment's results demonstrate the efficacy of the suggested approach in detecting false images.

Other biological signals such as heartbeat have been shown to be a reliable predictor for real video. Ciftci *et al.* [35] for example has designed a Generative Adversarial Network (GAN) based model that can detect the deepfake video source by analyzing the "heartbeat" of deep fakes. The proposed model starts by having several detector networks where the input to this model is the real video. Then, the pair of the realistic video and fake videos is assigned to another layer called registration, which extracts facial regions of interest (ROI) and the biological signals to create PPG cells. Here, PPG cells are spatiotemporal windows which contains multiple faces extracted using a face detector. The last layer is responsible for classifying the video as fake or real. The authors used several publicly available datasets to test their model. The result shows the models actives an accuracy of 97.3% in deepfake detection.

Prior research has shown that, in addition to biological signals, there is a close relationship between various audio-visual modalities of the same sample [36] [37] [38] [39] [40]. Mittal *et al.* [41] developed a deep learning framework for

detecting deepfake in multimedia materials. The primary goal of this model is to comprehend and examine the interaction of the audio (speech) and video (visual) modalities. To achieve that, the model uses a Siamese network-based architecture to simultaneously extract the speech and face modalities. To discriminate real and fake videos, the vectors representation for the video and audio of the sample are extracted using two modality embedding networks: OpenFace and pyAudioAnalysis respectively. Finally, a triplet loss function is used to calculate the similarity and identify the fake video and the real one. This approach is tested on two deepfake identification benchmark datasets, DeepfakeTIMIT dataset [42], and DFDC [43]. The models yield an accuracy of 96.6 percent on DF-TIMIT datasets and 84.4 percent on DFDC datasets, respectively.

2) Spatial and Temporal Features Analysis

Most current deepfake detection methods only use a single video frames [44]. In fact, video manipulation can be carried out on multiple frame-level features. Recently, many researches have shown that analyzing the temporal sequence between frames can successfully help to discriminate the real video or the fake one. In this paper [8], the authors introduced a temporally-aware model to detect deepfake videos. The model first employs a convolutional neural network (CNN) for frame features extraction. Afterwards, these features are passed to LSTM layer to analysis a temporal sequence for face manipulation between frames. Finally, a softmax function is used to classify the video as either real or fake. **Figure 4** illustrates the architecture of the model. For the evaluation, a collection of 600 videos was collected from multiple websites. The experimental results show the effectiveness of this model for deepfake video detection.

Based on the previous version of Cycle-GAN [45], Bansal *et al.* [46] introduced a new approach called, Recycle-GAN, which uses conditional generative adversarial networks to merge spatial and temporal data. The evaluation results show that combining the spatial and temporal constraints can produce an effective output. Furthermore, Sabir *et al.* [47] also propose a new approach based on recurrent convolutional network. The approach consists of two analysis stages: face processing stage followed by face manipulation detection. In the processing, face cropping and alignment is extracted using Spatial Transformer Network (STN). Then, the output from the previous stages is passed for face manipulation detection using the recurrent convolutional network, where the temporal information across frames is analyzed. See **Figure 5**. The approach is evaluated in a

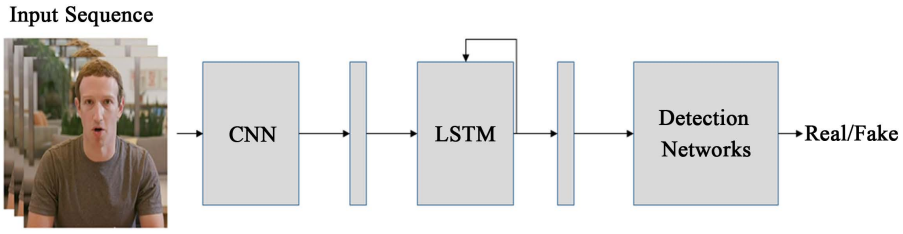

Figure 4. Convolutional neural network for spatial and temporal features analysis.

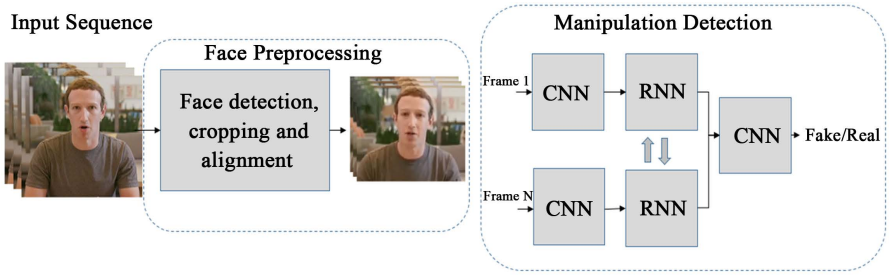

Figure 5. The proposed method is a two-step process. The first step is for face detection, cropping and alignment. The second step is for manipulation detection.

public available dataset FaceForensics++ [48]. The result shows state-of-the-art performance compared with the existing models.

4. Available Public Dataset

In this section, we present seven public datasets used for deepfake detection namely: 1) FFHQ; 2) 100K-Faces; 3) DFFD; 4) CASIA-WebFace; 5) VGGFace2 and (6) The eye-blinking dataset; 7) DeepfakeTIMIT. Table 1 contains information about these datasets, including download links.

4.1. Flickr-Faces-HQ, FFHQ

Karras *et al.* [49] introduced a dataset for human faces (Flickr-Faces-HQ, FFHQ). The dataset FFHQ contains a collection of 70,000 face images with a high-quality resolution generated by generative adversarial networks (GAN). The images were collected from Flicker platform and contain images with a variety of accessories such as eyeglasses, sunglasses, hats, etc. According to author, a pre-processing step was done in the dataset to prune the set, and remove noises from photos.

4.2. 100K-Faces

100K-Faces [50] is a well-known publicly available dataset which includes 100,000 unique human images generated using StyleGAN [49]. StyleGAN was applied on a large dataset consists of more than 29,000 images gathered from 69 different models, generating photos with a flat background.

4.3. Fake Face Dataset (DFFD)

Recently, another dataset called Diverse Fake Face Dataset (DFFD) was introduced by Dang *et al.* [51]. DFFD contains 100,000 and 200,000 fake images generated by adopting respective state-of-the-art methods (ProGAN and StyleGAN models). The dataset includes approximately 47.7 percent male photographs, 52.3 percent female images, and most of the samples range in age from 21 to 50 years old.

4.4. CASIA-WebFace

Dong *et al.* [52] presented a database called CASIA-WebFace that includes about

Table 1. Available public dataset.

Dataset	Type	Link
FFHQ	Images	https://github.com/NVlabs/stylegan
100K-Faces	Images	https://generated.photos/
DFFD	Images	https://github.com/NVlabs/ffhq-dataset
CASIA-WebFace	Images	https://paperswithcode.com/dataset/casia-webface
VGGFace2	Videos	https://www.tensorflow.org/datasets/catalog/vgg_face2
The Eye-Blinking	Videos	http://www.cs.albany.edu/%E2%88%BClsw/downloads.html
DeepfakeTIMIT	Videos	https://www.idiap.ch/en/dataset/deepfaketimit

10,000 subjects and 500,000 images. This dataset was first crawled from IMDB website which contains 10,575 of a well-known actors and actresses of IMDB. Then the photos of those celebrities are extracted using clustering methods.

4.5. VGGFace2

A large-scale face dataset called VGGFace2 was introduced by Cao *et al.* [53]. This database contains over three million face photographs from over nine thousand different subjects, with an average of more than 300 images per subject. Images were gathered from the Google engine which has a wide range of information such as ethnicity, illumination, age, and occupation (e.g., actors, athletes, and politicians).

4.6. The Eye-Blinking Dataset

The current available dataset has not been designed to deal with the eye-blinking detection. Li *et al.* [34] released the eye-blinking datasets which specially designed for this purpose. This dataset consists of 50 interviews and videos for each person for each person that lasts approximate thirty seconds with one eye blinking happened at least one time. Using their own tools, the author then tags the left and right eye states for each video clip. The details and description of the dataset is available at http://www.cs.albany.edu/%E2%88%BClsw/downloads.html .

4.7. DeepfakeTIMIT

DeepfakeTIMIT is a dataset of videos released by Korshunov *et al.* [42] using the database contains a collection of swapped faces videos generated using the GAN-based approach. The dataset was produced with a lower quality model with 64 × 64 size and a higher quality model with 128 × 128 input/output size. Each non-real video collection contains 32 subjects. The author created ten fictitious videos for each subject.

5. Challenges and Open Issues

The massive availability of applications and tools that create deepfake images and videos lead to large numbers of deepfake images and videos generated every

day. It has become a great challenge for academic researchers when studying and analyzing deepfake images and videos. One of the most important challenges facing the researchers is the lack of high-quality dataset. The current deep learning methods are facing a scalability issue. On other words, the deepfake methods uses fragmented data sets [32] to detect face swapping. However, applying their models in large scale datasets may produce unacceptable results. Another challenge the researchers should consider is the scalability issue. Second, the future studies require focusing on providing more robust and scalable models that can be applicable to any large and high-quality datasets. In addition to the previous challenges, it's well-known that deep learning models often require large dataset for the training step in order to produce good results, which, regrettably, are not freely accessible or need permission from social media providers. Fourth, the rapid development of deepfake GAN models can also bring a new challenge where unseen types of generated images and video may not be discovered by the current deep learning models. Hence, all these challenges discussed above can show the great demand to develop robust and scalable deep learning models to detect the fake images and videos.

6. Conclusion and Future Directions

Deepfake had become popular due to the massive availability of images and videos in social contents. This is particularly important nowadays because the tools for making deepfakes are becoming more accessible, and social media sites will easily allowing people to distribute and share such fake contents. Deep learning methods have received a lot of interest in a variety of areas. Recently, various deep learning-based methods have been proposed to address this issue and successfully detect fake images and videos. In this paper, we first discuss the current applications and tools that have been widely used to create fake images and videos. Then, we have reviewed current deepfake methods and divided them in this paper into two major techniques: image detection techniques and video detection techniques. We provided a detailed description of the current deepfake methods in terms of architecture, tool and performance. We also highlighted the publicly accessible datasets used by the science community, categorizing them by dataset sort, source, and method. Finally, we have also discussed the current challenges and provide insights into future research on deepfake detection using deep learning.

Although deep learning has shown a remarkable performance in deepfakes detection, the quality of deepfake has been increasing. Hence, the current deep learning methods need to improve as well to successfully identify fake videos and images. In addition, for the current deep learning methods, there is not a clear method to know the number of layers needed and which architecture is appropriate for deepfake detection. Another area of investigation is the incorporation of identification of deepfake detection methods into social media platform in order to improve their effectiveness in coping with the pervasive effects of deepfakes and reduce its impacts.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Nataraj, L., *et al.* (2019) Detecting GAN Generated Fake Images Using Co-Occurrence Matrices. *Electronic Imaging*, **2019**, 532-1-532-7. https://doi.org/10.2352/ISSN.2470-1173.2019.5.MWSF-532

[2] Wang, S.-Y., Wang, O., Zhang, R., Owens, A. and Efros, A.A. (2020) CNN-Generated Images Are Surprisingly Easy to Spot... for Now. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 8695-8704. https://doi.org/10.1109/CVPR42600.2020.00872

[3] Hsu, C.-C., Lee, C.-Y. and Zhuang, Y.-X. (2018) Learning to Detect Fake Face Images in the Wild. 2018 *IEEE International Symposium on Computer, Consumer and Control* (*IS3C*), Taichung, 6-8 December 2018, 388-391. https://doi.org/10.1109/IS3C.2018.00104

[4] Vaccari, C. and Chadwick, A. (2020) Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, **6**, 1-13. https://doi.org/10.1177/2056305120903408

[5] Mirza, M. and Osindero, S. (2014) Conditional Generative Adversarial Nets.

[6] Kwok, A.O. and Koh, S.G. (2020) Deepfake: A Social Construction of Technology Perspective. *Current Issues in Tourism*, 1-5. https://doi.org/10.1080/13683500.2020.1738357

[7] Westerlund, M. (2019) The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, **9**, 40-53. https://doi.org/10.22215/timreview/1282

[8] Güera, D. and Delp, E.J. (2018) Deepfake Video Detection Using Recurrent Neural Networks. 2018 15*th IEEE International Conference on Advanced Video and Signal Based Surveillance* (*AVSS*), Auckland, 27-30 November 2018, 1-6. https://doi.org/10.1109/AVSS.2018.8639163

[9] Li, Y. and Lyu, S. (2018) Exposing Deepfake Videos by Detecting Face Warping Artifacts.

[10] Yang, X., Li, Y. and Lyu, S. (2019) Exposing Deep Fakes Using Inconsistent Head Poses. 2019 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Brighton, 12-17 May 2019, 8261-8265. https://doi.org/10.1109/ICASSP.2019.8683164

[11] Marra, F., Gragnaniello, D., Cozzolino, D. and Verdoliva, L. (2018) Detection of Gan-Generated Fake Images over Social Networks. 2018 *IEEE Conference on Multimedia Information Processing and Retrieval* (*MIPR*), Miami, 10-12 April 2018, 384-389. https://doi.org/10.1109/MIPR.2018.00084

[12] Grekousis, G. (2019) Artificial Neural Networks and Deep Learning in Urban Geography: A Systematic Review and Meta-Analysis. *Computers, Environment and Urban Systems*, **74**, 244-256. https://doi.org/10.1016/j.compenvurbsys.2018.10.008

[13] Hopfield, J.J. (1982) Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, **79**, 2554-2558. https://doi.org/10.1073/pnas.79.8.2554

[14] Pouyanfar, S., *et al.* (2018) A Survey on Deep Learning: Algorithms, Techniques,

and Applications. *ACM Computing Surveys* (*CSUR*), **51**, 1-36. https://doi.org/10.1145/3234150

[15] Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y. (2016) Deep Learning (No. 2). MIT Press, Cambridge.

[16] Elman, J.L. (1990) Finding Structure in Time. *Cognitive Science*, **14**, 179-211. https://doi.org/10.1207/s15516709cog1402_1

[17] Bengio, Y., Simard, P. and Frasconi, P. (1994) Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Transactions on Neural Networks*, **5**, 157-166. https://doi.org/10.1109/72.279181

[18] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[19] Schuster, M. and Paliwal, K.K. (1997) Bidirectional Recurrent Neural Networks. *IEEE Transactions on Signal Processing*, **45**, 2673-2681. https://doi.org/10.1109/78.650093

[20] Faceswap: Deepfakes Software for All. https://github.com/deepfakes/faceswap

[21] FakeApp 2.2.0. https://www.malavida.com/en/soft/fakeapp

[22] Keras-VGGFace: VGGFace Implementation with Keras Framework. https://github.com/rcmalli/keras-vggface

[23] CycleGAN. https://junyanz.github.io/CycleGAN/

[24] Tariq, S., Lee, S., Kim, H., Shin, Y. and Woo, S.S. (2018) Detecting Both Machine and Human Created Fake Face Images in the Wild. *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security*, Toronto, 15 October 2018, 81-87. https://doi.org/10.1145/3267357.3267367

[25] Li, H., Li, B., Tan, S. and Huang, J. (2018) Detection of Deep Network Generated Images Using Disparities in Color Components.

[26] Do, N.-T., Na, I.-S. and Kim, S.-H. (2018) Forensics Face Detection from GANS Using Convolutional Neural Network. ISITC.

[27] Xuan, X., Peng, B., Wang, W. and Dong, J. (2019) On the Generalization of GAN Image Forensics. In: *Chinese Conference on Biometric Recognition*, Springer, Berlin, 134-141. https://doi.org/10.1007/978-3-030-31456-9_15

[28] Liu, F., Jiao, L. and Tang, X. (2019) Task-Oriented GAN for PolSAR Image Classification and Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 2707-2719. https://doi.org/10.1109/TNNLS.2018.2885799

[29] Zhou, P., Han, X., Morariu, V.I. and Davis, L.S. (2017) Two-Stream Neural Networks for Tampered Face Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Honolulu, 21-26 July 2017, 1831-1839. https://doi.org/10.1109/CVPRW.2017.229

[30] Hsu, C.-C., Zhuang, Y.-X. and Lee, C.-Y. (2020) Deep Fake Image Detection Based on Pairwise Learning. *Applied Sciences*, **10**, 370. https://doi.org/10.3390/app10010370

[31] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016) Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 2818-2826. https://doi.org/10.1109/CVPR.2016.308

[32] Nguyen, T.T., Nguyen, C.M., Nguyen, D.T., Nguyen, D.T. and Nahavandi, S. (2019) Deep Learning for Deepfakes Creation and Detection. Vol. 1.

[33] Afchar, D., Nozick, V., Yamagishi, J. and Echizen, I. (2018) Mesonet: A Compact

Facial Video Forgery Detection Network. 2018 *IEEE International Workshop on Information Forensics and Security* (*WIFS*), Hong Kong, 11-13 December 2018, 1-7. https://doi.org/10.1109/WIFS.2018.8630761

[34] Li, Y., Chang, M.-C. and Lyu, S. (2018) In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking. 2018 *IEEE International Workshop on Information Forensics and Security* (*WIFS*), Hong Kong, 11-13 December 2018, 1-7. https://doi.org/10.1109/WIFS.2018.8630787

[35] Ciftci, U.A., Demir, I. and Yin, L. (2020) How Do the Hearts of Deep Fakes Beat? Deep Fake Source Detection via Interpreting Residuals with Biological Signals. 2020 *IEEE International Joint Conference on Biometrics* (*IJCB*), Houston, 28 September-1 October 2020, 1-10. https://doi.org/10.1109/IJCB48548.2020.9304909

[36] Ciftci, U.A., Demir, I. and Yin, L. (2020) FakeCatcher: Detection of Synthetic Portrait Videos Using Biological Signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1 p. https://doi.org/10.1109/TPAMI.2020.3009287

[37] Shan, C., Gong, S. and McOwan, P.W. (2007) Beyond Facial Expressions: Learning Human Emotion from Body Gestures. *Proceedings of the British Machine Vision Conference* 2007, Coventry, 10-13 September 2007, 1-10. https://doi.org/10.5244/C.21.43

[38] Baltrušaitis, T., Ahuja, C. and Morency, L.-P. (2018) Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**, 423-443. https://doi.org/10.1109/TPAMI.2018.2798607

[39] Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C. and Foxe, J.J. (2007) Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. *Cerebral Cortex*, **17**, 1147-1153. https://doi.org/10.1093/cercor/bhl024

[40] Sanders, D.A. and Goodrich, S.J. (1971) The Relative Contribution of Visual and Auditory Components of Speech to Speech Intelligibility as a Function of Three Conditions of Frequency Distortion. *Journal of Speech and Hearing Research*, **14**, 154-159. https://doi.org/10.1044/jshr.1401.154

[41] Mittal, T., Bhattacharya, U., Chandra, R., Bera, A. and Manocha, D. (2020) Emotions Don't Lie: An Audio-Visual Deepfake Detection Method Using Affective Cues. *Proceedings of the* 28*th ACM International Conference on Multimedia*, Seattle, 12-16 October 2020, 2823-2832. https://doi.org/10.1145/3394171.3413570

[42] Korshunov, P. and Marcel, S. (2018) Deepfakes: A New Threat to Face Recognition? Assessment and Detection.

[43] Dolhansky, B., Howes, R., Pflaum, B., Baram, N. and Ferrer, C.C. (2019) The Deepfake Detection Challenge (DFDC) Preview Dataset.

[44] de Lima, O., Franklin, S., Basu, S., Karwoski, B. and George, A. (2020) Deepfake Detection Using Spatiotemporal Convolutional Networks.

[45] Zhu, J.-Y., Park, T., Isola, P. and Efros, A.A. (2017) Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2223-2232. https://doi.org/10.1109/ICCV.2017.244

[46] Bansal, A., Ma, S., Ramanan, D. and Sheikh, Y. (2018) Recycle-GAN: Unsupervised Video Retargeting. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Glasgow, 23-28 August 2018, 119-135. https://doi.org/10.1007/978-3-030-01228-1_8

[47] Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I. and Natarajan, P. (2019) Recurrent Convolutional Strategies for Face Manipulation Detection in Videos.

CVPR Workshops.

[48] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Nießner, M. (2019) Faceforensics++: Learning to Detect Manipulated Facial Images. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27-28 October 2019, 1-11. https://doi.org/10.1109/ICCV.2019.00009

[49] Karras, T., Laine, S. and Aila, T. (2019) A Style-Based Generator Architecture for Generative Adversarial Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 4401-4410. https://doi.org/10.1109/CVPR.2019.00453

[50] 100,000 Faces Generated by AI, 2018. https://generated.photos

[51] Dang, H., Liu, F., Stehouwer, J., Liu, X. and Jain, A.K. (2020) On the Detection of Digital Face Manipulation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 5781-5790. https://doi.org/10.1109/CVPR42600.2020.00582

[52] Yi, D., Lei, Z., Liao, S. and Li, S.Z. (2014) Learning Face Representation from Scratch.

[53] Cao, Q., Shen, L., Xie, W., Parkhi, O.M. and Zisserman, A. (2018) Vggface2: A Dataset for Recognising Faces across Pose and Age. 2018 13*th IEEE International Conference on Automatic Face & Gesture Recognition* (*FG* 2018), Xi'an, 15-19 May 2018, 67-74. https://doi.org/10.1109/FG.2018.00020