**DEEPFAKE: A BREIF SURVEY**

SEMINAR REPORT

Submitted by

**NIRANJAN N**

**16EC057**

To

*The University of Sri Siddhartha Academy of Higher Education*

*(SSAHE) -Tumkur*

*In Partial Fulfillment Of The Requirement For The Award Of B.E. Degree In*

*Electronics and Communication Engineering*



# Department of Electronics and Communication Engineering

Sri Siddhartha Institute of Technology-Tumkur

March 2020- 21

# Department of Electronics and Communication Engineering

Sri Siddhartha Institute of Technology,Tumkur



## **CERTIFICATE**

Certified that this report entitled ***‘DEEPFAKE: A SURVEY’*** is the report of Technical Seminar presented by **Niranjan N**, **16EC057** during **2020-2021** in partial fulfilment of the requirement for the award of the Degree of Bachelor of Engineering in Electronics and Communication of the University of the Sri Siddhartha Academy of Higher Education (SSAHE) Tumkur.

**Ravisimha B N Sharada Guptha M N**

**Assistant Professor Assistant Professor**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | |  | |  | |  | |  | |
|  |  |  |  |  | |  | |  | |  | |  | |

**Department of ECE** **Department of ECE**

**Dr. M Z Kurian**

**Head of the Department , ECE**

Name and Signature of HOD

## **1.ABSTRACT**

Advances in Artificial Intelligence and Image Processing are changing the way people

interacts with digital images and video. Widespread mobile apps like FACEAPP make use of the most advanced Generative Adversarial Networks (GAN) to produce extreme transformations on human face photos such gender swap, aging, etc. The results are utterly realistic and extremely easy to be exploited even for non-experienced users. This kind of media object took the name of Deepfake and raised a new challenge

in the multimedia forensics field: the Deepfake detection challenge. Indeed, discriminating a Deepfake from a real image could be a difficult task even for human eyes but recent works are trying to apply the same technology used for generating images for discriminating them with preliminary good results but with many limitations: employed Convolutional Neural Networks are not so robust, demonstrate to be specific to the context and tend to extract semantics from images. In this paper, a new approach aimed to extract a Deepfake fingerprint from images is proposed. The method is based on the Expectation-Maximization

algorithm trained to detect and extract a fingerprint that represents the Convolutional Traces (CT) left by GANs during image generation. The CT demonstrates to have high discriminative power achieving better results than state-of-the-art in the Deepfake detection task also proving to be robust to different attacks.

Achieving an overall classification accuracy of over 98%, considering Deepfakes from 10 different GAN architectures not only involved in images of faces, the CT demonstrates to be reliable and without any dependence on image semantic. Finally, tests carried out on Deepfakes generated by FACEAPP achieving 93% of accuracy in the fake detection task, demonstrated the effectiveness of the proposed technique on a real-case scenario.

*Index Terms*—Digital image forensics, video forensics, deep learning, Deepfakes.

# 2. LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| Figure No | Title | Page No |
| 1 | Examples of Deepfake manipulations from YouTube. | 8 |
| 2 | Examples of image and video manipulations carried out using deep learning methods | 12 |
| 3 | A Deepfake creation model using two encoder-decoder pairs | 13 |
| 4 | A two-step process for face manipulation detection | 16 |
| 5 | A Deepfake detection method using convolutional pairs | 16 |

# 3. CONTENTS

|  |  |  |
| --- | --- | --- |
| Chapter No | Title | Page No. |
| 1 | Introduction | 6 |
| 2 | Literature Review | 9 |
| 3 | Methodology | 11 |
| 4 | Application | 17 |
| 5 | Conclusion | 19 |
| 6 | Reference | 20 |

1. INTRODUCTION

Fake multimedia has become a central problem in the last few years, especially after the advent of the so called *Deepfakes*, i.e., fake media manipulated with the help of powerful and easy-to-use deep learning tools,[[1]](#footnote-1) like autoencoders (AE) or generative adversarial networks (GAN). With this technology, creating realistic manipulated media assets may be very easy, provided one can access large amounts of data. Applications include photography, video-games, virtual reality, and may soon expand to movie productions. The very same technology, however, can also be used for malicious purposes, like creating fake porn videos to blackmail people, or building fake-news campaigns to manipulate the public opinion. In the long run, it may also reduce trust in journalism, including serious and reliable sources. Figure 1 shows some popular deepfakes circulating on the internet. These fakes are easy to spot since they were generated for fun and involve well-known actors and politicians in unlikely situations. In addition, on the web it is usually possible to retrieve both the original and the manipulated version, removing any doubt about authenticity. However, verifying digital integrity becomes much more difficult if the video portrays a less known person and only the manipulated version is publicly available. This scenario takes place, for example, if the attacker films a new video on his own, with a collaborative actor whose face is eventually replaced by the face of the targeted person. Governmental bodies, enforcement agencies, the news industry, and also the man in the street are becoming acutely aware of the potential menace carried by such a technology. The scientific community is asked to develop reliable tools for automatically detecting fake multimedia.

Actually, this is not a new problem. Image manipulation has been carried out since photography was born, and powerful image/video editing tools, such as Photoshop R , After Effects Pro , or the open source software GIMP, have been around for a long time. Using such conventional tools images can be easily modified, obtaining realistic results that can fool even a careful observer. Figure 2 shows some examples of skillfully manipulated media, both imagesand videos, that have been disseminated on the Internet in recent years to spread false news. In fact, research in multimedia forensics has been going on for at least 15 years and is receiving ever growing attention, not only from the academy, but also from major information technology (IT) companies and funding agencies. In 2016, the Defense Advanced Research Projects Agency (DARPA) of the U.S. Department of Defense launched the large-scale Media Forensic initiative (MediFor)to foster research on media integrity, with important outcomes in terms of methods and reference datasets.

Following the MediFor taxonomy, digital media verification should look for physical integrity, digital integrity, and semantic integrity. In the literature, several methods have been proposed, which expose physical inconsistencies, concerning for example shadows or illumination or perspective . Modern sophisticated manipulations, however, are more and more effective in avoiding such pitfalls and methods which test digital integrity are by far more widespread and represent the current state of the art. Indeed, each image or video is characterized by a number of features, which depend on the different phases of its digital history: from the very same acquisition process, to the internal camera processing (e.g. demosaicing, compression), to all external processing and editing operations Digital manipulations tend to modify such features, leaving a trail of clues which, although invisible to the eye, can be exploited by pixel-level analysis tools. Instead, semantic integrity is violated when the media asset under analysis conveys information which is not coherent with the context or with evidence coming from correlated sources. For example, when objects are copy-pasted from images available on the web, several near-identical copies can be detected suggesting a possible manipulation. Moreover, by identifying the connections among the various versions of the same asset, it is possible to build its manipulation history (image and video phylogeny).

Despite the continuous research efforts and the numerous forensic tools developed in the past, the advent of deep learning is changing the rules of the game and asking multimedia forensics for new and timely solutions. This phenomenon is also causing a strong acceleration in multimedia forensics research, which often relies itself on deep learning. There have already been several reviews on this topic, but the advent of so many new methods in recent years calls for continuous updating. Hence, beyond reviewing the conventional media forensics approaches, a special attention will be devoted to deep learning-based approaches and to the strategies designed to fight deepfakes. Moreover, it will be assumed that the attacker modifies the metadata to make it useless, otherwise it would provide precious information towards authenticity verification both for images and videos . On the other hand, it is worth noting that metadata is routinely canceled when media assets are uploaded on a social network. The analysis will be restricted to passive methods and visual data-based solutions. However, it is important to underline that active methods may be very effective in ensuring the digital integrity of media assets, and there is a growing interest on these authentication methods to protect media content . In past decades, a large body of research was produced on digital watermarking . There is now much interest in blockchain technology , in cryptography, and even new active methods have been proposed to ensure the integrity of digital media or to protect individuals from becoming the victims of AI attacks.

The review starts with a brief analysis of the most effective manipulation methods proposed in recent years . Then, integrity verification methods are described, beginning with conventional approaches, then moving to deep learning-based approaches, and finishing with specific deepfake detection methods. In Section, a discussion of the state of multimedia forensics and its perspectives after the advent of deep learning is carried out. A list of the datasets most widespread in the field is presented . Then, the further major themes of counterforensics and fusion are considered. Finally, future research directions are outlined and conclusions are drawn.



Fig. 1. Examples of deepfake manipulations from YouTube. Top: manipulated videos; bottom: original videos.

## 2.LITERATURE REVIEW

**2.1** The spread of smartphones with high quality digital cameras in combination with easy access to a myriad of software apps for recording, editing and sharing videos and digital images in combination with deep learning AI platforms has spawned a new phenomenon of faking videos known as Deepfake. We design and implement a deep-fake detection model with mouth features (DFT-MF), using deep learning approach to detect Deepfake videos by isolating, analyzing and verifying lip/mouth movement. Experiments conducted against datasets that contain both fake and real videos showed favorable classification performance for DFT-MF model especially when compared with other work in this area**.**

**2.2**  With the rapid progress of recent years, techniques that generate and manipulate multimedia content can now provide a very advanced level of realism. The boundary between real and synthetic media has become very thin. On the one hand, this opens the door to a series of exciting applications in different fields such as creative arts, advertising, film production, video games. On the other hand, it poses enormous security threats. Software packages freely available on the web allow any individual, without special skills, to create very realistic fake images and videos. These can be used to manipulate public opinion during elections, commit fraud, discredit or blackmail people. Therefore, there is an urgent need for automated tools capable of detecting false multimedia content and avoiding the spread of dangerous false information. This review paper aims to present an analysis of the methods for visual media integrity verification, that is, the detection of manipulated images and videos. Special emphasis will be placed on the emerging phenomenon of deepfakes, fake media created through deep learning tools, and on modern data-driven forensic methods to fight them. The analysis will help highlight the limits of current forensic tools, the most relevant issues, the upcoming challenges, and suggest future directions for research.

**2.3** Deep learning has been successfully applied to solve various complex problems ranging from big data analytics to computer vision and human-level control. Deep learning advances however have also been employed to create software that can cause threats to privacy, democracy and national security. Oneof those deep learning-powered applications recently emerged is “deepfake”. Deepfake algorithms can create fake images and videos that humans cannot distinguish them from authentic ones. The proposal of technologies that can automatically detect and assess the integrity of digital visual media is therefore indispensable. This paper presents a survey of algorithms used to create deepfakes and, more importantly, methods proposed to detect deepfakes in the literature to date. We present extensive discussions on challenges, research trends and directions related to deepfake technologies. By reviewing the background of deepfakes and state-of-the-art deepfake detection methods, this study provides a comprehensive overview of deepfake techniques and facilitates the development of new and more robust methods to deal with the increasingly challenging deepfakes.

**2.4** DeepFake using Generative Adversarial Networks (GANs) tampered videos reveals a new challenge in today’s life. With the inception of GANs, generating highquality fake videos becomes much easier and in a very realistic manner. Therefore, the development of efficient tools that can automatically detect these fake videos is of paramount importance. The proposed DeepFake detection method takes the advantage of the fact that current DeepFake generation

algorithms cannot generate face images with varied resolutions, it is only able to generate new faces with a limited size and resolution, a further distortion and blur is needed to match and

fit the fake face with the background and surrounding context in the source video. This transformation causes exclusive blur inconsistency between the generated face and its background in the outcome DeepFake videos, in turn, these artifacts can be effectively spotted by examining the edge pixels in the wavelet domain of the faces in each frame compared to the rest of the frame. A blur inconsistency detection scheme relied on the type of edge and the analysis of its sharpness using Haar wavelet transform as shown in this paper, by using this feature, it can determine if the face region in a video has been blurred or not and to what extent it has been blurred. Thus will lead to the detection of DeepFake videos. The effectiveness of the proposed scheme is demonstrated in the experimental results where the “UADFV” dataset has been used for the evaluation, a very successful detection rate with more than 90.5% was gained.

**3. METHODOLOGY**

3.1 FAKE CONTENT GENERATION

There are many ways to manipulate visual content, and new methods are proposed by the day. This Section will briefly review some of the most widespread and promising among them. Very common operations are adding, replicating or removing objects, as in the examples . A new object can be inserted by copying it from a different image (splicing), or from the same image (copy-move). Instead, an existing object can be deleted by extending the background to cover it (inpainting) like in the popular exemplar-based inpainting . All these tasks are easily accomplished with widespread image editing packages. Then, some suitable post-processing, like resizing, rotation or color adjustment, can also be applied to

better fit the object to the scene, both to improve the visual appearance and to guarantee coherent perspective and scale. In recent years, however, the same results are achieved, with better semantic consistency, through advanced computer graphics (CG) approaches and deep learning (see Figure 4, last column). Manipulations that do not require sophisticated artificial intelligence (AI) tools are sometimes referred to as “shallowfakes” or “cheapfakes” (which is somewhat ironic, since generating deepfakes is often less expensive). Nonetheless, their impact in distorting reality can be very high. For example, by removing, inserting or cloning entire groups of frames one can completely change the meaning of a video.

A simple frame-rate reduction was recently used to let Nancy Pelosi, Speaker of the U.S. House of Representatives, appear as drunk or confused.

Besides these “traditional” manipulations, concerning specific areas of the image or video, deep learning and computer graphics are now offering a large number of new ones. Sometimes, media assets are synthesized completely from scratch. To this end, autoencoders and generative adversarial networks allowed to develop successful solutions especially for face synthesis, where a high level of photo-realism has been achieved. It is also possible to generate a completely synthetic image or video using a segmentation map as input

. Image synthesis is also achievable using only a sketch or a text description . Likewise, the face of a person can be animated based on an audio input sequence

More often, the manipulation modifies existing images or videos. A well-known example is style transfer which allows to change the style of a painting, switch oranges to apples, or reproduce an image in a different season. Major efforts have been devoted to manipulating faces, for their high semantic value, and for the many possible applications. Methods have been proposed to change the expression of a face to transfer the expression from a source to a target actor , or to swap faces . Recently, it has been shown that effective face manipulation is feasible even without a huge amount of training photos of the targeted person. It is even possible to animate the face of a still portrait and express various types of emotions . Beyond faces, some recent work addressed motion transfer: the target person dances following the movements transferred from a source dancer . In Figur some examples of such manipulations are presented. One can easily observe how realistic they appear and the variety of possible automatic editing tools available nowadays.

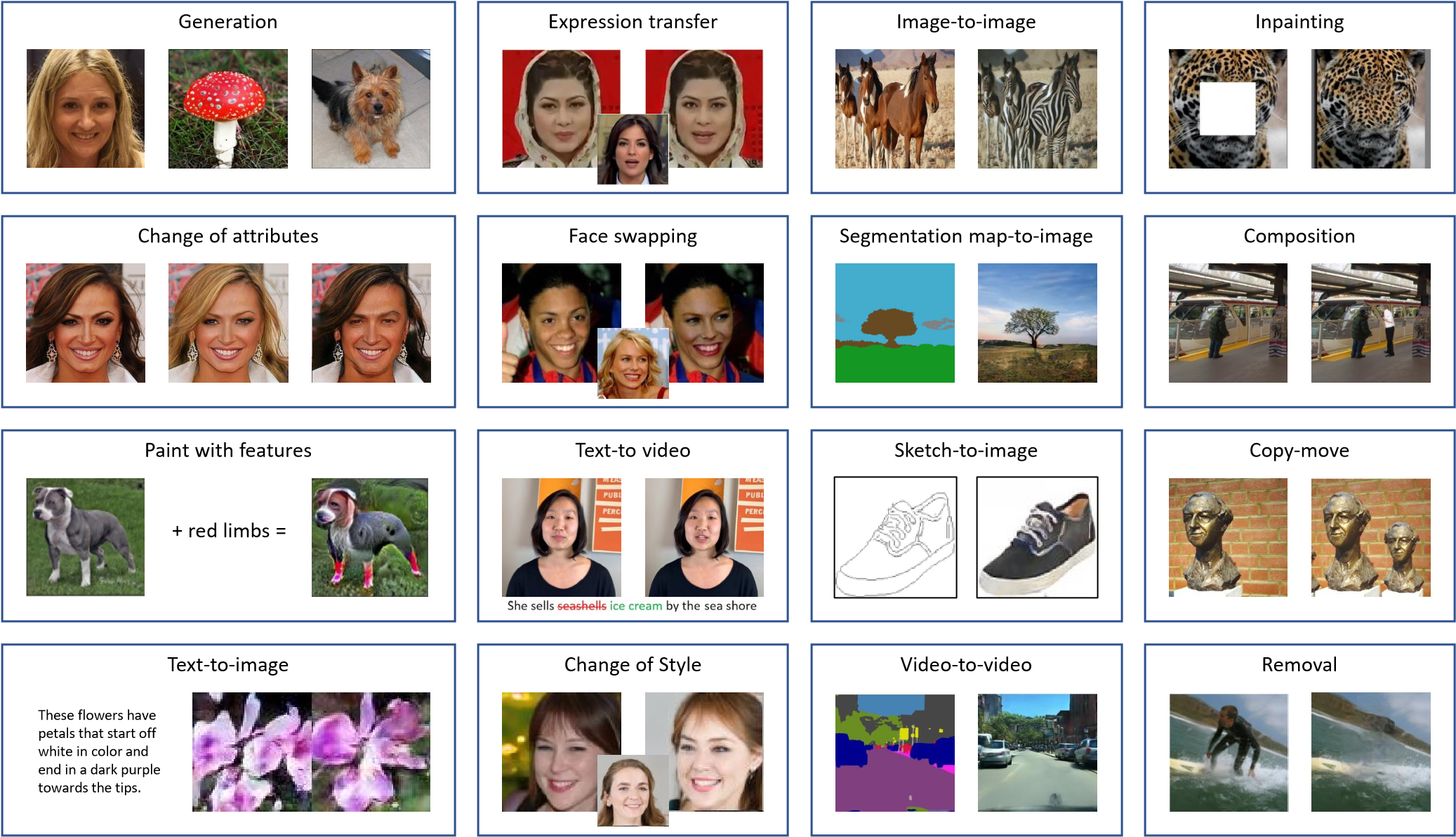


Fig. 2. Examples of image and video manipulations carried out using deep learning methods

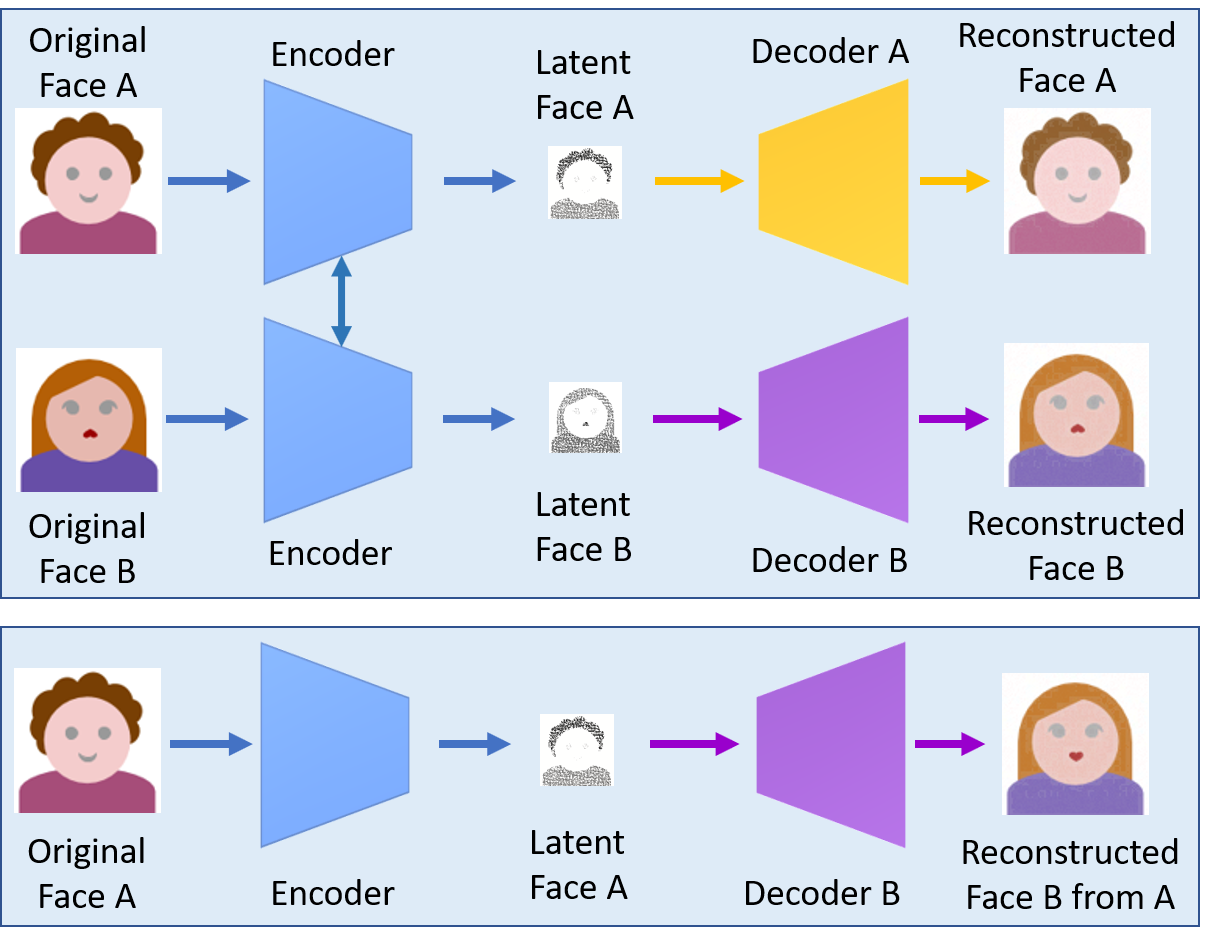


Fig.3 A deepfake creation model using two encoder-decoder pairs

3.2 DEEPFAKE DETECTION

Manipulations carried out using deep learning tools have become a major issue nowadays. Especially alarming is how easily false media can be created by relatively unskilled individuals, provided sufficient data are available. Currently, deep learning-based attacks focus on two main targets, the creation of fully synthetic images by means of GAN-based architectures, and the manipulation of faces in videos aimed at changing identities or semantics. Accordingly, this Section is divided in two subsections which describe forensic methods for these two approaches.

*A. Fake Image Detection*

Face swapping has a number of compelling applications in video compositing, transfiguration in portraits, and especially in identity protection as it can replace faces in photographs by ones from a collection of stock images. However, it is also one of the techniques that cyber attackers employ to penetrate identification or authentication systems to gain illegitimate access. The use of deep learning such as CNN and GAN has made swapped face images more challenging for forensics models as it can preserve pose, facial expression and lighting of the photographs . Zhang used the bag of words method to extract a set of compact features and fed it into various classifiers such as, random forest and multi-layer perceptrons for discriminating swapped face images from the genuine. Among deep learninggenerated images, those synthesised by GAN models are probably most difficult to detect as they are realistic and highquality based on GAN’s capability to learn distribution of the complex input data and generate new outputs with similar input distribution.

Most works on detection of GAN generated images however do not consider the generalization capability of the detection models although the development of GAN is ongoing, and many new extensions of GAN are frequently introduced. Xuan used an image preprocessing step, e.g. Gaussian blur and Gaussian noise, to remove low level high frequency clues of GAN images. This increases the pixel level statistical similarity between real images and fake images and requires the forensic classifier to learn more intrinsic and meaningful features, which has better generalization capability than previous image forensics methods or image steganalysis networks.

On the other hand, Agarwal and Varshney cast the GAN-based deepfake detection as a hypothesis testing problem where a statistical framework was introduced using the information-theoretic study of authentication . The minimum distance between distributions of legitimate images and images generated by a particular GAN is defined, namely the oracle error. The analytic results show that this distance increases when the GAN is less accurate, and in this case, it is easier to detect deepfakes. In case of high-resolution image inputs, an extremely accurate GAN is required to generate fake images that are hard to detect.

Recently, Hsu introduced a two-phase deep learning method for detection of deepfake images. The first phase is a feature extractor based on the common fake feature network (CFFN) where the Siamese network architecture presented in is used. The CFFN encompasses several dense units with each unit including different numbers of dense blocks to improve the representative capability for the fake images. The number of dense units is three or five depending on the validation data being face or general images, and the number of channels in each unit is varied up to a few hundreds. Discriminative features between the fake and real images, i.e. pairwise information, are extracted through CFFN learning process. These features are then fed into the second phase, which is a small CNN concatenated to the last convolutional layer of CFFN to distinguish deceptive images from genuine. The proposed method is validated for both fake face and fake general image detection. On the one hand, the face data set is obtained from CelebA , containing 10,177 identities and 202,599 aligned face images of various poses and background clutter. Five GAN variants are used to generate fake images with size of 64x64, including deep convolutional GAN (DCGAN) , Wasserstein GAN (WGAN) , WGAN with gradient penalty (WGAN-GP) , least squares GAN , and progressive growth of GAN (PGGAN) . A total of 385,198 training images and 10,000 test images of both real and fake ones are obtained for validating the proposed method. On the other hand, the general data set is extracted from the ILSVRC12 . The large scale GAN training model for high fidelity natural image synthesis (BIGGAN) , self-attention GAN and spectral normalization GAN are used to generate fake images with size of 128x128. The training set consists of 600,000 fake and real images whilst the test set includes 10,000 images of both types. Experimental results show the superior performance of the proposed method against its competing methods such as those introduced later.

*B. Fake Video Detection*

Most image detection methods cannot be used for videos because of the strong degradation of the frame data after video compression. Furthermore, videos have temporal characteristics that are varied among sets of frames and thus challenging for methods designed to detect only still fake images. This subsection focuses on deepfake video detection methods and categorizes them into two groups: methods that employ temporal features and those that explore visual artifacts within frames.

Video manipulation is carried out on a frame-by-frame basis so that low level artifacts produced by face manipulations are believed to further manifest themselves as temporal artifacts with inconsistencies across frames. A recurrent convolutional model (RCN) was proposed based on the integration of the convolutional network DenseNet and the gated recurrent unit cells [85] to exploit temporal discrepancies across frames .

On the other hand, the use of a physiological signal, eye blinking, to detect deepfakes was proposed in based on the observation that a person in deepfakes has a lot less frequent blinking than that in untampered videos. A healthy adult human would normally blink somewhere between 2 to 10 seconds, and each blink would take 0.1 and 0.4 seconds. Deepfake algorithms, however, often use face images available online for training, which normally show people with open eyes, i.e. very few images published on the internet show people with closed eyes. Thus, without having access to images of people blinking, deepfake algorithms do not have the capability to generate fake faces that can blink normally. In other words, blinking rates in deepfakes are much lower than those in normal videos. To discriminate real and fake videos, Li first decompose the videos into frames where face regions and then eye areas are extracted based on six eye landmarks. After few steps of pre-processing such as aligning faces, extracting and scaling the bounding boxes of eye landmark points to create new sequences of frames, these cropped eye area sequences are distributed into long-term recurrent convolutional networks (LRCN) for dynamic state prediction. The LRCN consists of a feature extractor based on CNN, a sequence learning based on long short term memory (LSTM), and a state prediction based on a fully connected layer to predict probability of eye open and close state. The eye blinking shows strong temporal dependencies and thus the implementation of LSTM helps to capture these temporal patterns effectively. The blinking rate is calculated based on the prediction results where a blink is defined as a peak above the threshold of 0.5 with duration less than 7 frames. This method is evaluated on a data set collected from the web consisting of 49 interview and presentation videos and their corresponding fake videos generated by the deepfake algorithms. The experimental results indicate promising performance of the proposed method in detecting fake videos, which can be further improved by considering dynamic pattern of blinking, e.g. highly frequent blinking may also be a sign of tampering.

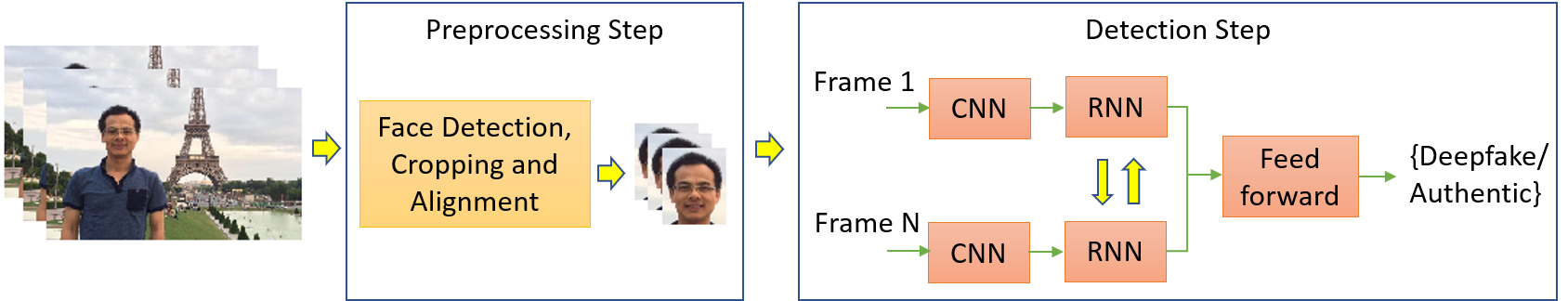


Fig. 4. A two-step process for face manipulation detection

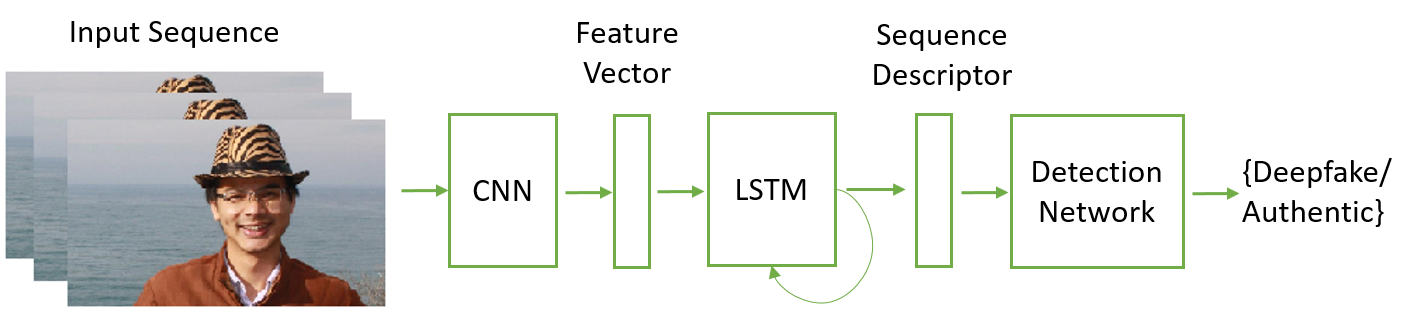


Fig. 5. A deepfake detection method using convolutional neural network

**4. APPLICATIONS**

AI-Generated Synthetic media, also known as deepfakes, have many positive use cases. AI-Generated Synthetic Media, aka Deepfakes, advances have clear benefits in certain areas, such as accessibility, education, film production, criminal forensics, and artistic expressions.

## **Accessibility:** Artificial intelligence can build tools to hear, see, and soon with Artificial General Intelligence (AGI), reason with increasing accuracy. AI-Generated Synthetic media can help make the accessibility tools smarter and, in some cases, even affordable and personalizable, which can help people augment their agency and gain independence. Technology, specifically AI-based tools, can bring accessible solutions to all.

## Microsoft’s Seeing.ai and Google’s Lookout leverage AI for recognition and synthetic voice to narrate objects, people, and the world. AI-Generated synthetic media can power personalized assistive navigation apps for pedestrian travel.

## For the Amyotrophic Lateral Sclerosis (ALS) patients’ synthetic voice is very empowering as the horrible disease impacts the ability to speak, communicate, and motor skills. Imagine being able to talk in your own voice with your loved ones even after losing the ability to speak. Team Gleason, which provides technology, equipment, and services for people living with ALS, is working with technology companies to enable and develop AI-Generated synthetic media scenarios for people living with ALS (Lou Gehrig’s Disease).

## **Education:** Deepfake technology facilitates numerous possibilities in the education domain. Schools and teachers have been using media, audio, video in the classroom for quite some time. Deepfakes can help an educator to deliver innovative lessons that are far more engaging than traditional visual and media formats.

AI-Generated synthetic media can bring historical figures back to life for a more engaging and interactive classroom. A synthetic video of re-enactments or voice and video of a historical figure may have more impact, engagement, and will be a better learning tool. For example, JFK’s resolution to end the cold was speech, which was never delivered, was recreated using synthetic voice with his voice and speaking style will clearly get students to learn about the issue in a creative way.

## **Art:** For many decades, Hollywood has used high-end CGI, VFX, and SFX technologies to create artificial but believable worlds for compelling storytelling. In the 1994’s movie, Forest Gump, the protagonist meets JFK and other historical figures. The [creation](https://snippetofhistory.wordpress.com/portfolio/forrest-gump-with-president-john-f-kennedy-1994-movie/) of the scenario and effect was accomplished using CGI and different techniques with millions of dollars. These days sophisticated CGI and VFX technologies are used in movies to generate synthetic media for telling a captivating story.

and technicians can work together in industrial products together with [Remote Assist](https://dynamics.microsoft.com/en-us/mixed-reality/remote-assist/), an mixed reality collaboration tool.

AI technologies can also be used to enhance and improve the resolution of low-resolution images. These enhancing techniques to build deepfakes are super useful for older media or media created by low-resolution lenses. Google published a paper to [create](https://arxiv.org/pdf/2003.02365v1.pdf) high-resolution images using Latent Adversarial Generator. Microsoft Research published an article to enhance and [fix imperfections](https://www.microsoft.com/en-us/research/publication/bringing-old-photos-back-to-life/) of the degraded photos using triplet Domain Translation Networks.

Synthetic data can empower medical researchers to develop new ways of treating diseases without actual patient data. They can use AI-Generated synthetic data for the training model to yield similar results. NVIDIA, MGH & BWH Center for Clinical Data Science and the Mayo Clinic showed how they used GANs–algorithms that iterate and improve by competing against each other–to create synthetic brain MRI images with tumors[[4](https://arxiv.org/pdf/2006.07397.pdf)]. By training algorithms on synthetic medical images and combining them with just 10% real images, they became just as good at spotting tumors as algorithms trained only on real images.

A vast amount of training data is required by deep learning algorithms to create an efficient model and even to develop deepfake detection models. Since there are not many real-life deepfakes, companies like Facebook had to create synthetic media and deepfakes for the Facebook Deepfake detection challenge[[5](https://arxiv.org/pdf/2006.07397.pdf)]

**5. CONCLUSION**

Fifteen years ago multimedia forensics was a niche field of practical interest only for a restricted set of players involved in law enforcement, intelligence, private investigations. Both attacks and defences had an artisan flavour, and required painstaking work and dedication

Artificial intelligence has largely changed these rules. High quality fakes now seem to come out from an assembly line calling for an extraordinary effort on part of both scientists and policymakers. In fact, today’s multimedia forensics is in full development, major agencies are funding large research initiatives, and scientists from many different field are contributing actively, with fast advances in ideas and tools.

It is difficult to forecast whether such efforts will be able to ensure information integrity in the future, or some forms of active protection will become necessary. This is an arms race, and one part is no smarter than the other. For the present time, a large arsenal of tools is being developed, and knowing them, the principles on which they rely, and their scope of application is a prerequisite to protect institutions and ordinary people.

**6. REFERENCES**

1. H. Farid, “Image forgery detection,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, 2009.
2. ——, *Photo Forensics*. The MIT Press, 2016.
3. M. Johnson and H. Farid, “Exposing digital forgeries in complex lighting environments,” *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 450–461, 2007.
4. E. Kee, J. O’Brien, and H. Farid, “Exposing photo manipulation with inconsistent shadows,” *ACM Transactions on Graphics*, vol. 32, no. 3, pp. 28–58, 2013.
5. T. de Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. Rocha, “Exposing digital image forgeries by illumination color classification,” *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 7, pp. 1182–1194, 2013.
6. A. Piva, “An overview on image forensics,” *ISRN Signal Processing*, pp. 1–22, 2012.
7. Y. Wu, W. Abd-Almageed, and P. Natarajan, “Deep matching and validation network: An end-to-end solution to constrained image splicing localization and detection,” in *ACM International Conference on Multimedia*, 2017, pp. 1480–1502.
8. Y. Lui, X. Zhu, X. Zhao, and Y. Cao, “Adversarial learning for constrained image splicing detection and localization based on atrous convolution,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 10, pp. 2551–2566, 2019.
9. L. Kennedy and S.-F. Chang, “Internet image archaeology: automatically tracing the manipulation history of photographs on the web,” in *ACM international conference on Multimedia*, 2008, pp. 349–358.
10. T.-C. Wang, M.-Y. Liu, A. Tao, G. Liu, J. Kautz, and B. Catanzaro, “Few-shot Video-to-Video Synthesis,” in *Neural Information Processing Systems*, 2019.
11. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
12. S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, “Generative adversarial text-to-image synthesis,” in *International Conference on Machine Learning*, 2016.
13. P.-W. Wu, Y.-J. Lin, C.-H. Chang, E. Chang, and S.-W. Liao, “RelGAN: Multi-Domain Image-to-Image Translation via Relative Attributes,” in *International Conference on Computer Vision*, 2019.
14. L. Engstrom, A. Ilyas, S. Santurkar, D. Tsipras, B. Tran, and A. Madry, “Adversarial robustness as a prior for learned representations,” *arXiv preprint arXiv:1906.00945v2*, 2019.
15. J. Y. Zhu, T. Park, P. Isola, and A. Efros, “Unpaired image-toimage translation using cycle-consistent adversarial networks,” in *IEEE International Conference on Computer Vision*, 2017.
16. Y. Nirkin, I. Masi, A. T. Tran, T. Hassner, and G. Medioni, “On face segmentation, face swapping, and face perception,” in *IEEE Conference on Automatic Face and Gesture Recognition*, 2018.
17. O. Fried, A. Tewari, M. Z. abd A. Finkelstein, E. Shechtman, D. Goldman, K. Genova, Z. Jin, C. Theobalt, and M. Agrawala, “Text-based editing of talking-head video,” *ACM Transactions on Graphics*, vol. 38, no. 4, 2019.
18. C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, “PatchMatch: A randomized correspondence algorithm for structural image editing,” *ACM Transactions on Graphics*, vol. 28, no. 3, 2009.
19. H. Huang, P. Yu, and C. Wang, “An introduction to image synthesis with generative adversarial nets,” *arXiv:1803.04469v2*, 2018.

1. [↑](#footnote-ref-1)