Status, Challenges, and Future Views of DeepFake Techniques and Datasets

Ahmed S Abdulreda, Ahmed J Obaid
Faculty of Computer Science and Mathematics, University of Kufa, Iraq
ahmeds.albudairi@student.uokufa.edu.iq , ahmedj.aljanaby@uokufa.edu.iq

Article Info
Page Number: 225 – 233
Publication Issue:
Vol 71 No. 2 (2022)

Article History
Article Received*: 30 December 2021*
Revised*: 30 January 2022*
Accepted*: 15 March 2022*
Publication*: 07 April 2022*

Abstract
Deep counterfeiting is a term that has become popular in recent years and carries many risks in violation of privacy and other dangers, which were addressed by several research studies that used different methods and methods in counterfeiting and anti-counterfeiting detection methods, we will address in this paper one of the methods that help in detecting image manipulation specifically in a modified manner on the GAN algorithm and we will call it ADVGAN and it includes the process of selecting and configuring a specific database Several initial operations and training by learning the machine and relying on the proposed system in detecting image manipulation with excellent efficiency.

Keywords: fraud, counterfeiting, deep counterfeiting, Deep faking , faking.

1. Introduction

This paper discusses the strategies that deal with deepfaking and its types, the data used in manipulation, and the methods for detecting counterfeiting. As for the databases, the operations that will be performed on them are filtering and selecting the celebrity dataset (Celeb-DF), preparing the database, and then selecting the images, cutting frames, cleaning, partitioning, merging and training. The focus is on the data related to the celebrity image database, which will be the focus of the paper, and the proposed system is applied to this data set, and the most important topic of the study (deep fake) focus on the most important type is the deepfake of facial images. And what are its types that have four main types which are identity fraud, complete forgery or barter of expression, and manipulation of characteristics, and each type has many aspects. In terms of anti-counterfeiting, we review the most common methods currently used and the most important forgery detection method that our paper will focus on, the method is to train a neural network on a set of images called two types, the real and the fake. Then the characteristics of dataset were explained in terms of most of the features that this data set possesses. The forgery methods used are the topic that occupies the bulk of the paper [1] .

2. Common Deepfake Techniques

2.1. Visual content: This means the use of “Deepfakes” technology to create pictures and videos, which in turn includes several sections, including:

2.1.1. Full face swapping: Using encryption / decoding and decryption algorithms because the two sides are different, the first algorithm is programmed to recover the second person's face. To

switch between the two sides, the instructions are provided with a decoding algorithm with the image data encoded from the other side [2] Figure (1).

Figure1 Full face swapping using Encoder/Decoder

2.1.2. Facial Manipulation: Modify his expressions and lip sync using generative adversarial networks. This method uses two artificial intelligence algorithms, where random data is entered into the first algorithm known as the generation algorithm to convert it into an image. Then this fake photo was added and it is part of a series of real photos of some celebrities. For example, it is fed into a second algorithm known as the discrimination algorithm. At first, the generated images do not look like faces figure(3),(4), except to repeat the process several times and make adjustments based on performance feedback, which leads to better figure(5) "highlighting" algorithms and the creation of new images. Generator After making enough cycles and observations, the algorithm starts producing realistic faces that are not real people figure(5) at all. Generative adversarial networks, also referred to as adversarial networks, are a type of machine learning network developed by Ian Goodfellow and colleagues in 2014.

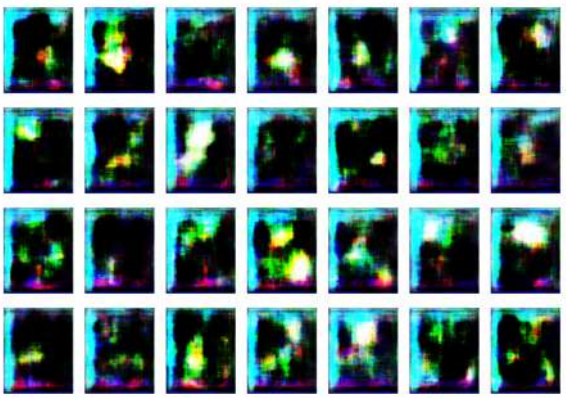
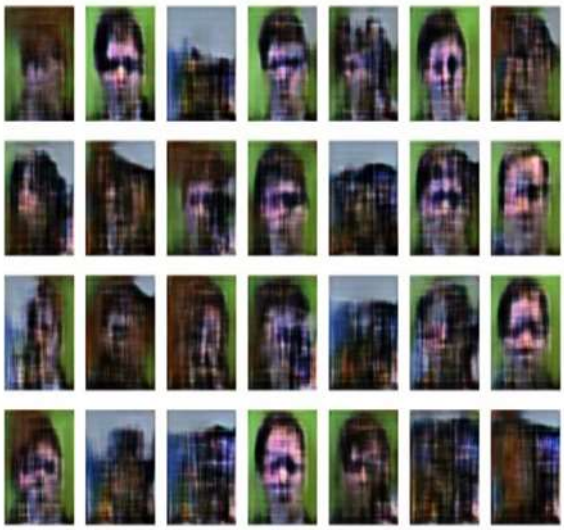


Figure 4 1st Stage of face generation

Figure 3 several cycles of generation (Some facial features are starting to appear)

Figure 1 Imagined by a GAN (generative adversarial network) StyleGAN2 (Dec 2019) - Karras et al. and Nvidia



1.1.1. Figure 2 After several cycles of generation (better)

2.1.3. Features Manipulation: Forgery in some facial features was the change in the eyes or nose and other characteristics of the face. Several incidents of this kind occurred, similar to the first case of the birth of this technology by obtaining videos of a football celebrity, appearing as a real video on social networking sites, while he was doing household activities, talking about certain topics, and moving around in the outside yard of the house, the video gained millions of views in a short time [2], [3].

2.1.4. Expression Swapping: It is meant to transfer a specific expression from one person to another who did not and common expressions such as sadness, joy, smile, frown and other facial features. As a result, among the applications of this technology are several programs on smart phone platforms that move the face of the person who did not move, and move it in the position that the phone user wants, either with ready movements and prepared templates such as dances or by tracking the movement of the owner of the phone by opening the phone's camera [4].

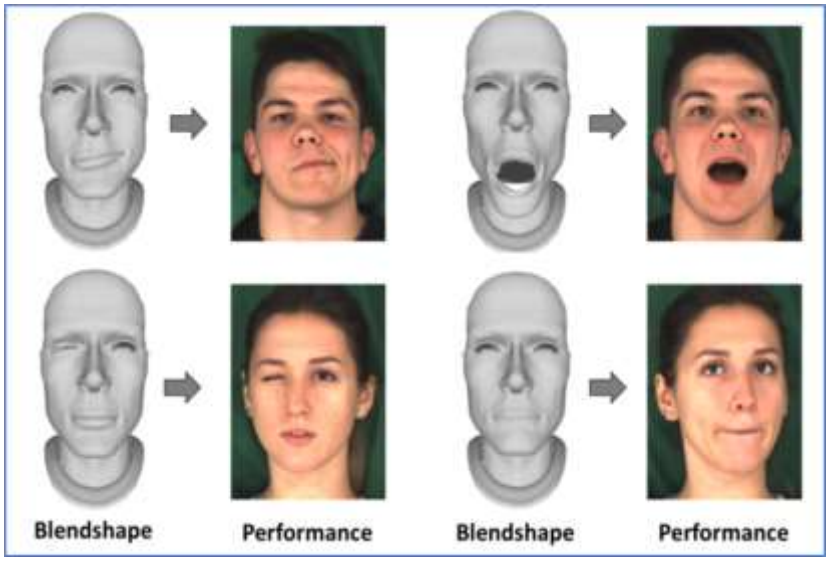


Figure 5 Expression Swapping sourse and target

2.2. Audio content manipulation

Basically it means stabilizing and modifying the voice either by creating an audio file containing fake speech in the same voice as the person but not actually saying it, or by controlling the person's tone of voice to show an unrealistic feeling or behavior. Deepfakes poses a huge risk of tampering with facts and discrediting by broadcasting these messages on various media channels without revealing their source. Although these techniques are complex for non-AI specialists, the Internet provides many tools and applications that allow anyone to create deep fake content in real time on their phones and computers.

Deepfake technology has been used in many ways to target people in all walks of life. Not only has it been used to create fake photos and videos of celebrities and politicians, but this technology has also been used to defraud companies and steal their money [5].

3. Common Deepfake Datasets

Among the important system requirements, is the selection of a suitable database to work on. The online databases, on which studies were conducted previously, vary from 2014 until writing this paper, and there are many data sets of different sizes and different contents, Table (5) shows some of the data sets used or training and research [6],[7]. Table 1 example of Datasets [4],[7].

Table 1 deepfake datasets used for training

1^{St} Generation		
Database	Real videos	Fake Videos
UADFV 2018	49 (Youtube)	49 (Fake App)
Deepfake TIMIT(2018)	-	620 (faceswap-GAN)
FaceForensiecs++(2019)	1.000 (Youtube)	1000(faceswap) 1000(Deepfake)
2^{ND} Generation		
Deepfake Detection (2019)	363 (Actors)	3038 (Deepfake)
Celeb-DF (2019)	**890(youtube)**	**5639(deepfake)**
DFDC Preview (2019)	1131(Actors)	4119(Unknown)

In the proposed system, a specific dataset, which is a set of videos of the Dataset Celeb-DF, Celeb-DF is a challenging large-scale dataset for deep forensics, will be handled in the proposed system. It includes 890 original videos collected from YouTube with subjects of different ages, ethnic groups, and genders and 5639 matching DeepFake videos. The dataset can be obtained by submitting a request to the author via Google Form and agreeing to the Scientific Use Agreement only without commercial use. When approval is obtained, the institution will send the dataset link via email. It is a requirement that the researcher have an academic email and fill in some data related to the work project. Link to get the dataset:

(https://docs.google.com/forms/d/e/1FAIpQLScoXint8ndZXyJi2Rcy4MvDHkkZLyBFKN43lTeyiG88wrG0rA/viewform) Author Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu Title Celeb-DF: A Challenging Large-Scale Dataset for DeepFake Forensics Title Book IEEE Conference on Computer Vision and Patten Recognition (CVPR) 2020 dataset of celeb-DF. One of the reasons for choosing this database is that it is free, easy to access, short time to provide it to the researcher, and it is sufficient for the research requirements table 2 show some datasets use with face synthesis.

Table 2 Face Synthesis Available Datasets [2]

Database	Real images	Fake Images
100K-Generated -Images(2019)[8]	-	100000 (Style GAN)
100K-Faces (2019)[9]	-	100000 (Style GAN)
DFFD 2020[10]	-	100000 (Style GAN) 200000 (Pro GAN)
I FakeFaceDB (2020)[11]	-	250000 (Style GAN) 80000 (Pro GAN)

Researchers make many modifications to the database data before starting to use it to avoid making mistakes, avoiding deviations, and modifying the data and unifying it in one format, so choosing the appropriate rule saves the researcher some editing and pre-modification. The processing that this dataset needs in such a proposed system is to extract images from video clips and get a new set of data in the form of images to work on instead of the video. In the third chapter of this research, we will explain in detail many steps regarding the method of extracting images from the video [3], [12].

4. Benefits of working on the Celb-DF database

One of the features that benefit us by working on this dataset is the closed front image mode for people, which produces images with few defects with processing and training. And because we know that all the images are arranged in advance, we used the method of distinguishing faces through the Cascade classifier instead of the classifier that works with deep learning technology because the data is for faces in advance and there is no need to waste time in a complicated way. In order to gain additional time during the work of the system, we removed a feature from the workbook that we see as redundant in the system and consumes some time, which is the function of rotating images and searching for images tilted at several angles, because it consumes some unnecessary time [13]

The tool tilts the images in several angles to get the optimal and most appropriate face position, and because the database that we filtered in advance, the data is organized in the form of a front video clip that contains the scene on one character depicted from a front angle, we got rid of the complexity and waste of time in this tool and we stopped its work to gain some time and reduce time complexities. And get the desired results, which are close-up images of people's faces directly.

5. Current Challenges of Deepfake techniques

Challenging Scenario Augmentation to enhance the challenges posed by real-world face forgery detection and segmentation, applied various perturbations to better simulate contexts in natural

scenes, resulting in a test-challenge subset. Various augmented operators are divided into overarching groups.

5.1 Color manipulation: Hue change, saturation change, brightness change, histogram adjustment, contrast addition, grayscale conversion.

5.2 Edge manipulation: edge detection and alteration.

5.3 Block-wise distortion: color grouping, color pooling, color quantization, and pixelation.

5.4 Image corruption: elastic deformation, jigsaw distortion, JPEG compression, noise addition, and dropout.

5.5 Convolution mask transformation: Gaussian blurring, motion blurring, sharpening, and embossing.

5.6 External effect: fog, cloud, sun, frost, snow, and rain.

6. Deepfake Detection Methods

6.1 Two-stream :uses a two-stream CNN to achieve state-of-the-art performance in general-purpose image forgery detection. The underlying CNN is the GoogLeNet InceptionV3 model trained on the SwapMe dataset . We use it as a baseline to compare other dedicated DeepFake detection methods [6].

6.2 MesoNet: is a CNN-based DeepFake detection method targeting on the mesoscopic properties of images. The model is trained on unpublished DeepFake datasets collected by the authors. We valuate two variants of MesoNet, namely, Meso4 and MesoIncep- tion4. Meso4 uses conventional nvolutional layers, while MesoInception4 is based on the more sophisticated Inception modules [14].

6.3 HeadPose : detects DeepFake videos using the inconsistencies in the head poses of the synthesized videos, based on a SVM model on estimated 3D head orientations from each video. The SVM model in this method is trained on the UADFV dataset [6].

6.4 FWA : detects DeepFake videos using a ResNet-50 to expose the face warping artifacts introduced by the resizing and interpolation operations in the basic DeepFake maker algorithm. This model is trained on self-collected face images [15].

6.5 VA : is a recent DeepFake detection method based on capturing visual artifacts in the eyes, teeth and facial contours of the synthesized faces. There are two variants of this method: VA-MLP is based on a multilayer feedforward neural network classifier, and VA-LogReg uses a simpler logistic regression model. These models are trained on unpublished dataset, of which real images are cropped from CelebA dataset and the DeepFake videos are from YouTube [16].

6.6 Xception : corresponds to a DeepFake detection method based on the XceptionNet model trained on the FaceForensics++ dataset. There are three variants of Xception, namely, Xception-raw, Xception- c23 and Xception-c40: Xception-raw are trained on raw videos, while Xception-c23 and Xception-c40 are trained on H.264 videos with medium (23) and high degrees (40) of compression, respectively[17], [18].

6.7 Multi-task : is another recent DeepFake detection method that uses a CNN model to simultaneously detect manipulated images and segment manipulated areas as a multi-task learning problem. This model is trained on the FaceForensics dataset [7].

6.8 Capsule : uses capsule structures based on a VGG19 network as the backbone architecture for DeepFake classification. This model is trained on the FaceForensics++ dataset [8].

6.9 DSP-FWA : is a recently further improved method based on FWA, which includes a spatial

pyramid pooling (SPP) module to better handle the variations in the resolutions of the original target faces. This method is trained on self-collected face image s. A concise summary of the underlying model, source code, and training datasets of the DeepFake detection methods considered in our experiments is given in Table 3.

Table 3 Summary of compared DeepFake detection methods[10].

Methods	Model Type	Training Dataset	Repositories	Release Date
Two-stream	GoogLeNet InceptionV3	SwapMe	Unpublished code provided by the authors	2018.03
MesoNet	Designed CNN	Unpublished	https://github.com/DariusAf/MesoNet	2018.09
HeadPose	SVM	UADFV	https://bitbucket.org/ericyang3721/headpose_forensic/	2018.11
FWA	ResNet-50	Unpublished	https://github.com/danmohaha/CVPRW2019_Face_Artifacts	2018.11
VA-MLP	Designed CNN	Unpublished	https://github.com/FalkoMatern/Exploiting-Visual-Artifacts	2019.01
VA-LogReg	Logistic Regression	FaceForensics++	https://github.com/ondyari/FaceForensics	2019.01
Xception	XceptionNet	FaceForensics	https://github.com/nii-yamagishilab/ClassNSeg	2019.06
Multi-task	Designed CNN	FaceForensics++	https://github.com/nii-yamagishilab/Capsule-Forensics-v2	2019.10
Capsule	Designed CapsuleNet	Unpublished	https://github.com/danmohaha/DSP-FWA	2019.11

There is a clear discrepancy in the level of Accuracy of techniques in detecting counterfeiting. It can be seen in the Figure 6 that shows the technique used and the percentage of the Accuracy achieved figure (9) [10].

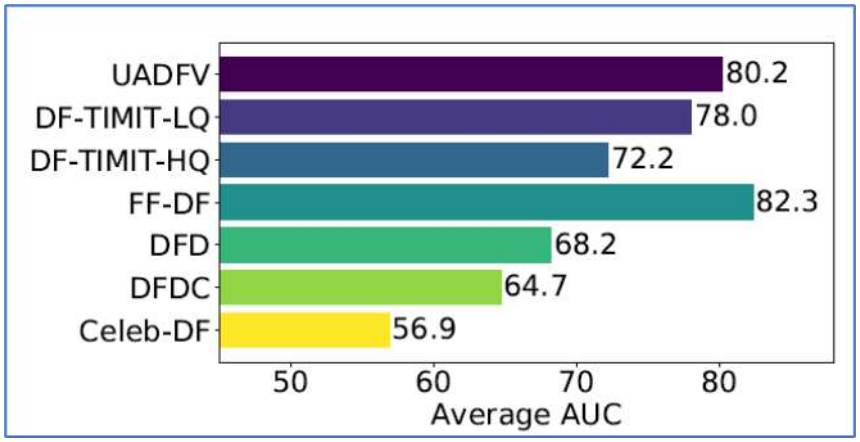


Figure 7 datasets used and there accuracy

7. Conclusion

This paper adopts how to detect deepfakes in faces by training a neural network of two main layers, one generative and the other critical, as in the generative adversarial network, but instead of generating images in the generator layer, named images are passed to this layer and the critical layer

depends The properties of these images to allow the tested images to pass through or not. One of the system requirements is to choose a dataset and make some modifications to the contents as initial steps before passing it on to the network and then conduct the training and testing phases that will take place in any proposed system. However, this paper list the most used deepfake Detection techniques upon well known Deepfake datasets.

Reference

[1] 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, doi: 10.22075/IJNAA.2022.5580.

[2] S. Sankaranarayanan, Y. Balaji, C. D. Castillo, and R. Chellappa, "Generate to Adapt: Aligning Domains Using Generative Adversarial Networks," 2018. doi: 10.1109/CVPR.2018.00887.

[3] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," Jan. 2020, [Online]. Available: http://arxiv.org/abs/2001.00179

[4] Emily Barrow, "https://www.regendus.com/best-face-swap-apps/," Aug. 2019.

[5] Jesse Damiani, "A Voice Deepfake Was Used To Scam A CEO Out Of $243,000," Sep. 03, 2019.

[6] B. Dolhansky *et al.*, "The DeepFake Detection Challenge (DFDC) Dataset," 2020, [Online]. Available: http://arxiv.org/abs/2006.07397

[7] 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," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 8789–8797, 2018, doi: 10.1109/CVPR.2018.00916.

[8] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 4396–4405, 2019, doi: 10.1109/CVPR.2019.00453.

[9] "100,000 Faces Generated by AI, 2018. [Online]. Available https".

[10] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, "On the Detection of Digital Face Manipulation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 5780–5789, 2020, doi: 10.1109/CVPR42600.2020.00582.

[11] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing obama: Learning lip sync from audio," in *ACM Transactions on Graphics*, 2017, vol. 36, no. 4. doi: 10.1145/3072959.3073640.

[12] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3204–3213, 2020, doi: 10.1109/CVPR42600.2020.00327.

[13] D. Chen, S. Ren, Y. Wei, X. Cao, and J. Sun, "LNCS 8694 - Joint Cascade Face Detection and Alignment," 2014.

[14] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A compact facial video forgery detection network," *10th IEEE International Workshop on Information Forensics and Security, WIFS 2018*, 2019, doi: 10.1109/WIFS.2018.8630761.

[15] M. Liu *et al.*, "STGAN: A unified selective transfer network for arbitrary image attribute editing," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 3668–3677, 2019, doi: 10.1109/CVPR.2019.00379.

[16] T. Jung, S. Kim, and K. Kim, "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," *IEEE Access*, vol. 8, pp. 83144–83154, 2020, doi: 10.1109/ACCESS.2020.2988660.

[17] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, "FaceForensics++: Learning to detect manipulated facial images," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-Octob, pp. 1–11, 2019, doi: 10.1109/ICCV.2019.00009.

[18] R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12665 LNCS, pp. 442–456, 2021, doi: 10.1007/978-3-030-68821-9_38.

[19] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Use of a Capsule Network to Detect Fake Images and Videos," 2019, [Online]. Available: http://arxiv.org/abs/1910.12467