2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon) | 979-8-3503-0541-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/SmartTechCon57526.2023.10391497

Deep fake Video Detection Using Unsupervised Learning Models: Review

B N Jyothi
Research Scholar, JNTUH
Department of Computer Science and Engineering
Vardhaman College of Engineering
Shamshabad, India.
jyothi.jo515@gmail.com

M. A. Jabbar
Department of CSM
Vardhaman College of Engineering
Shamshabad, India
jabbar.meerja@gmail.com

Abstract

The recent decade has been the most emerging for technical advancements in computer and communication fields. These advancements have made the task of automation much cheaper and available to every other individual at minimum or free of cost. Deep Fakes are one such advancement in the field of image processing. Deep fakes are images or videos developed using deep learning models to create fake images or videos. Although there are advantages of these deep fakes like visual training, powerful simulation, etc... There are also demerits or disadvantages highly possible with these technological advancements such as creating content that never existed, making a person speak something which one has not spoken at all. These mentioned possibilities lead to person identification theft, spreading of false or misinformation, etc... This paper is an attempt to investigate the proposed detection techniques for deep fakes. It focuses on unsupervised deep fake detection techniques and the role of ensemble models for detection.

Keywords: Deep Learning, Deep fakes, deep fake detection, unsupervised learning, an ensemble of model.

1. Introduction

Emerging technological advancements bring a few challenging outcomes like deep fakes. The Recent advancements in smart devices and network connectivity aid people in generating and sharing images and videos with ease. With the increased availability of images, videos, and sophisticated manipulation algorithms like deep generative models has made the manipulation of images easy and less time-consuming. These manipulations lead to problems like distrust, identity theft, etc... These manipulations portray the individuals doing or saying things they never said or done. These lead to situations like cyber bullying, fake news spread, etc [1].

Deep learning techniques are used to forge both images and videos. These manipulations are so convincing that it is difficult to identify them with the naked eye[2] Availability of packages like Keras, and tensor flow along with readily trained models make the task of generating deep fakes easier for even laypersons [3].

2. Origin of image manipulation

Photo manipulation has its roots in the 19th century. During the late 20th century, Photography evolved into the digital mode. Digital image processing since its inception has undergone major improvements. The availability of sophisticated computers at lower cost and improved network availability fueled the digital imaging field majorly [4].

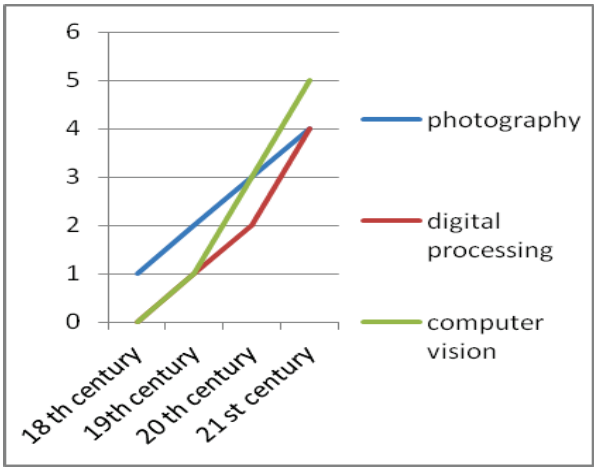

Fig: 1 Evolution of Image Manipulation

3. Deep fakes

Deep fakes are fake content created using deep learning techniques. These are so authentic that it is difficult to detect for the human eye Deep fakes have both positive and negative outcomes. The manipulations of the content such as images and videos are of multiple types such as reenactment, replacement, editing, and synthesis [5].

Reenactment: reenactment is used to manipulate the source in terms of, pose, gaze, body, mouth, etc…All these manipulations have their positive uses as well as negative impact on society when used by people with wrong intentions. These make the target person say or do what he/she has not done in reality. For example, using mouth reenactment to make a person say something controversial.

Authorized licensed use limited to: Manipal University. Downloaded on September 24,2024 at 09:18:45 UTC from IEEE Xplore. Restrictions apply.

Replacement: Replacement is transferring or swapping one person's face with another. This kind of deep fake is commonly used for revenge porn.

Editing: editing is majorly meant for entertainment purposes such as improving color, and hair, reducing age, etc...But these also have the possibility of defaming through the removal of attire, making a person look sick when they are healthy, etc.

Synthesis: This technique is used to create synthetic data such as TV show characters, some online fake characters, etc.

Although all these manipulations can harm a person's identity majority of research is centered around reenactment and replacement. The following sections summarize various deep fake creation models and detection models along with the deep learning models used by these models and various datasets used.

3.1 Impact of deep fakes:

Deep Fakes when used for educational purposes like recreating historical events, real-time simulations, etc.

In the following sections, we summarize a few of the deep fake creation and detection models and their respective accuracies.

4. Deep fake creation models

The table deep fake creation model summarizes various deep learning models used in the generation of deep fakes along with the dataset used. Please refer to Table 1 deep fake creation models.

5. Deep fake detection models

The table deep fake detection models summarize various deep learning models used in the generation of deep fakes along with the dataset used. Please refer to Table 2 deep fake creation models.

6. DL models in deep fake detection.

The majority of existing Detection models implements CNN models. The following figure illustrates the usage of various DL techniques.

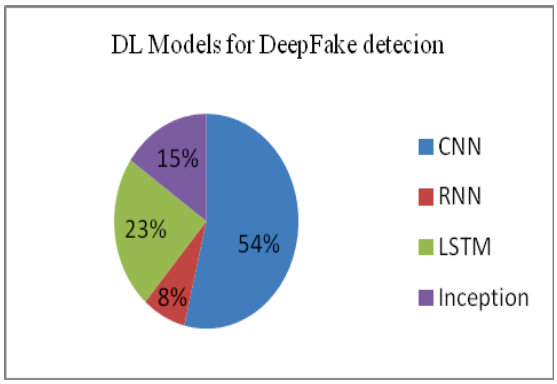

Table1. Deep Fake Creation Models

Ref no	Model name	DL technique used	Year of publication	Datasets used
1	Avatar Me++: [6]	GANFIT (Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction)	01 December,2022	Real Face DB
2	Face Shifter:[7]	Adaptive Attentional Denormalization layers + Heuristic Error Acknowledging Refinement Network (HEAR-Net)	15 September 2020	Face Forensics++
3	DeepFaceLab: [8]	S3FD for face detection, heatmap-based facial landmark algorithm 2DFAN and PRNet for face alignment and encoder and decoder for generalization	29 June 2021	Deep fakes, and Face2Face Face Forensics++ dataset.

Authorized licensed use limited to: Manipal University. Downloaded on September 24,2024 at 09:18:45 UTC from IEEE Xplore. Restrictions apply.

4	FSGAN:[9]	recurrent neural network (RNN) + VGG-19 CNNs for the perceptual loss	16 August 2019	VGGFace2 dataset, CelebA dataset, Forensics++
5	First Order Motion Model for Image Animation[10]	Monkey-Net	29 February 2020	VoxCeleb, Tai-Chi-HD, Fashion-Videos and MGif .
6	A Lip Sync Expert Is All You Need for Speech to Lip Generation[11]	Wav2Lip model + GAN	23 august 2020	ReSyncED, a Real-world Lip-Sync Evaluation Dataset
7	Face2Face: [12]	monocular facial reenactment in real-time	29 July 2020	
8	Exploring Deep fakes - Creation Techniques, Detection Strategies, and Emerging Challenges: [13]	Unsupervised method, Cycle GAN that can execute image-to-image conversion without the necessity for matched samples	March 2023	Over 600 fake videos generated via GAN, utilizing the open-source software Face swap-GANs.
9	Deep Insights of Deep fake Technology: A Review[14]	deep auto encoders with 2 uniform deep networks of 4 or 5 layers Face Swap, GANS, MTCNN, Deep Face Lab	1 May 2021	---
10	Deep fakes Creation and Detection Using Deep Learning[15]	Auto encoder-decoder pipelineESRGAN+ DFD Net	09 June 2021	Dataset of more than 5000 images

Table2. Deep Fake Detection Models

SNO	Model name	DL Technique used	Year of publication	Accuracy
1	Deep fake Video Detection System Using Deep Neural Networks[16]	RNN Residual Neural Network50, LSTM	2023	Real - 99% Fake - 50%
2	Deep fake Face Detection Using Deep Inception Net Learning Algorithm[17]	CNN Convolution Neural Network Inception Net	2023	87.40%
3	Deep Fake Detection Through Key Video Frame Extraction using GAN[18]	CNN Resnext50 & GAN LSTM Pre-trained model	2023	97.20%
4	An Efficient Deep Fake Video Detection Approach with Combination of Efficient Net and Xception Models Using Deep Learning[19]	Xception and EfficientNetB4	2022	83.17%

Authorized licensed use limited to: Manipal University. Downloaded on September 24,2024 at 09:18:45 UTC from IEEE Xplore. Restrictions apply.

5	Deep fake Video Detection Using Spatiotemporal Convolution Network and Photo Response Non-Uniformity[20]	Spatiotemporal Convolution Network & Photo-Response Non-Uniformity (PRNU)	2022	Real - 99.98% Fake - 83.45%
6	Forgery Detection Scheme of Deep Video Framerate Up-conversion Based on Dual-stream Multi-scale Spatial-temporal Representation[21]	CNN (MSDN) Convolution Neural Network for Microsoft Developer Network & FRUC	2022	-
7	Adversarial robust deep fake media detection using fused convolution neural network predictions[22]	Convolution Neural Network (CNN) models, VGG16, InceptionV3, and XceptionNet	2021	96.50%
8	Multi-Modal Deep Fake Detection Using Attention Based Ensemble Learning[23]	CNN ResNet-50, Xception	2020	89%
9	Detection of Real-world Fights in Surveillance Videos[24]	CNN, LSTM, SVM	2019	76%

7. Unsupervised deep fake detection models.

Unsupervised learning models are of significant help when available data is unlabeled. Supervised models require the data to be labeled. With a huge amount of raw data labeling, it might not be a practical solution.
We consider unsupervised learning using the frequency domain and spatial domain.

7.1 Frequency Domain: In [25] the authors have used Fast Fourier transform to decompose the spatial signals in preprocessing step followed by 2D and 1D spectrum alongside azimuthally average to represent the signal vectors which are given as input to the classifier while training.

Another approach [26] uses the Haar wavelet Transform for the generation of low-frequency feature maps. These feature maps and the grayscale of the input image are used to calculate the residual image from which they obtained the mid-high frequency maps. The obtained frequency maps are concatenated with RGB Domain input. The concatenated content is the input to the classification module where the pre-trained Xception net is used for classification. The overall accuracy obtained when compared with other methods is improved.

7.2 Spatial Domain

Commonality-based detection is the strategy used by Peipeng Yu et al [27] where multiple forgery feature extractors are used to supervise a common feature extractor. The proposed method works on the fact that each fake creation model leaves some traces in images which are common among multiple creation models.

Sheldon Fung et al [28] have proposed an unsupervised deep fake detection framework based on the principle of contrastive learning which when compared with the state-of-the-art supervised models outperforms other methods when tested against the same dataset and gives a comparable output when checked against the different data set.
L Zhang et al proposed a method of clustering using device fingerprint and photo response non uniformity which is a two-step clustering process.[29]

Cross-data set generalization is the major issue faced by many deep fake detection models as a result of the over fitting of deep learning models on source data sets. The model [30] proposes a back propagation-based domain adversarial neural network model where the preprocessing step works on extracting features that distinguish the target and source domain using transfer learning followed by back propagation-based training.

8. Ensemble of detection methods

Ensemble techniques have been used to increase the predictive performance of any given model through a combination of models rather than using a single model. The ensemble has been successfully used by various authors in deep fake detection.

In[31] Md. Shohel Rana and others have proposed a deep fakes stack where a few Deep learning models such as Dense Net, Xception Net, etc... Are used as base learners and a CNN model as a meta learner. This architecture achieves 99.65 % accuracy. Similarly, another method employs a technique where residual signals from multiple color spaces are considered to study the image representations, and the

Authorized licensed use limited to: Manipal University. Downloaded on September 24,2024 at 09:18:45 UTC from IEEE Xplore. Restrictions apply.

concatenated outcomes are fed into a random forest classifier this ensemble model achieves above 90 % of accuracy [32].

9. Discussions

We have tried understanding different manipulations images undergo while generating deep fakes. There is research going on and datasets available for the domain, but due to the dynamic nature of creation models detection models which are best classifying one kind of manipulation are failing to detect others.

Having said this the research done so far is majorly discussing the architectural aspects rather than technical insights into the work. So, future surveys can consider exploring technical aspects rather than architectural part. The paper also highlights a few of the research problems that are being identified during the survey.

10. Research Scope

This survey identifies a few research problems in deep fake detection

a) Availability of datasets with images manipulated by diverse generation models.

b) Generalization of proposed models across different data sets.

c) Reducing the overall training time required for the model to converge.

11. Conclusion

This work is an attempt to provide an in-depth understanding of Deep learning models' role in Deep fake creation and detection. We considered major research from 2019 to 2023, an attempt has been made to segregate and understand different creation models, the way they improved performance over time, and the reasons behind the failure of different detection models in the detection of deep fakes.

Apart from the above, the survey includes issues regarding Cross data set generalization of detection models, unsupervised models in improving model generalization, and an ensemble of deep learning models in deep fake detection. Overall, this survey provides an understanding of unsupervised learning models in deep fake detection, the accuracy obtained, and an ensemble of these models in further improvements in the detection of deep fakes. The improving technological advancements make creation models powerful making it difficult for the detection models difficult to cope with these creation models. This gives future scope for researchers to aim for more sophisticated detection models which adapt to the new data and provide better results

References

[1] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini, and S. Tubaro, "Video Face Manipulation Detection Through Ensemble of CNNs," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.07676

[2] "A Qualitative Survey on Deep Learning Based Deep fake Video Creation and Detection Method," *Aust. J. Eng. Innov. Technol.*, pp. 13–26, Feb. 2022, doi: 10.34104/ajeit.022.013026.

[3] M. Masood, M. Nawaz, K. M. Malik, A. Javed, and A. Irtaza, "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward."

[4] R. C. Gonzalez, R. E. Woods, and P. Prentice Hall, "Digital Image Processing Third Edition Pearson International Edition prepared by Pearson Education."

[5] Y. Mirsky and W. Lee, "The Creation and Detection of Deepfakes: A Survey," Apr. 2020, doi: 10.1145/3425780.

[6] A. Lattas, S. Moschoglou, S. Ploumpis, B. Gecer, A. Ghosh, and S. Zafeiriou, "AvatarMe++: Facial Shape and BRDF Inference with Photorealistic Rendering-Aware GANs," Dec. 2021, doi: 10.1109/TPAMI.2021.3125598.

[7] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen, "FaceShifter: Towards High Fidelity And Occlusion Aware Face Swapping," Dec. 2019, [Online]. Available: http://arxiv.org/abs/1912.13457

[8] I. Perov *et al.*, "DeepFaceLab: Integrated, flexible and extensible face-swapping framework," May 2020, [Online]. Available: http://arxiv.org/abs/2005.05535

[9] S. Chaudhary, R. Saifi, N. Chauhan, and R. Agarwal, "A Comparative Analysis of Deep Fake Techniques," in *Proceedings - 2021 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 300–303. doi: 10.1109/ICAC3N53548.2021.9725392.

[10] M. B. Priya and J. F. Daniel, "First Order Motion Model for Image Animation and Deep Fake Detection: Using Deep Learning," in *2022 International Conference on Computer Communication and Informatics, ICCCI 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICCCI54379.2022.9740969.

[11] K. R. Prajwal, R. Mukhopadhyay, V. P. Namboodiri, and C. V. Jawahar, "A Lip Sync Expert Is All You Need for Speech to Lip Generation in the Wild," in *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, Oct. 2020, pp. 484–492. doi: 10.1145/3394171.3413532.

[12] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2Face: Real-time Face Capture and Reenactment of RGB Videos," Jul. 2020,

Authorized licensed use limited to: Manipal University. Downloaded on September 24,2024 at 09:18:45 UTC from IEEE Xplore. Restrictions apply.

[Online]. Available: http://arxiv.org/abs/2007.14808

[13] J. Gadgilwar, K. Rahangdale, O. Jaiswal, P. Asare, P. Adekar, and P. L. Bitla, "Exploring Deepfakes - Creation Techniques, Detection Strategies, and Emerging Challenges: A Survey," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 3, pp. 1491–1495, Mar. 2023, doi: 10.22214/ijraset.2023.49681.

[14] B. Uddin Mahmud and A. Sharmin, "D Deep Insights of Deepfake Technology : A Review."

[15] H. A. Khalil and S. A. Maged, "Deepfakes Creation and Detection Using Deep Learning," in *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference, MIUCC 2021*, Institute of Electrical and Electronics Engineers Inc., May 2021, pp. 24–27. doi: 10.1109/MIUCC52538.2021.9447642.

[16] S. R. B. R, P. Kumar Pareek, B. S, and G. G, "Deepfake Video Detection System Using Deep Neural Networks," in *2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)*, IEEE, Feb. 2023, pp. 1–6. doi: 10.1109/ICICACS57338.2023.10099618.

[17] P. Theerthagiri and G. B. Nagaladinne, "Deepfake Face Detection Using Deep InceptionNet Learning Algorithm," in *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/SCEECS57921.2023.10063128.

[18] S. Lalitha and K. Sooda, "DeepFake Detection Through Key Video Frame Extraction using GAN," in *International Conference on Automation, Computing and Renewable Systems, ICACRS 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 859–863. doi: 10.1109/ICACRS55517.2022.10029095.

[19] S. Atas, I. Ilhan, and M. Karakse, "An Efficient Deepfake Video Detection Approach with Combination of EfficientNet and Xception Models Using Deep Learning," in *2022 26th International Conference on Information Technology, IT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/IT54280.2022.9743542.

[20] S. J. Pipin, R. Purba, and M. F. Pasha, "Deepfake Video Detection Using Spatiotemporal Convolutional Network and Photo Response Non Uniformity," in *ICOSNIKOM 2022 - 2022 IEEE International Conference of Computer Science and Information Technology: Boundary Free: Preparing Indonesia for Metaverse Society*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICOSNIKOM56551.2022.10034890.

[21] Q. Gu, X. Ding, D. Zhang, and C. Yang, "Forgery Detection Scheme of Deep Video Frame-rate Up-conversion Based on Dual-stream Multi-scale Spatial-temporal Representation," in *Proceedings - 2022 IEEE 21st International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 733–738. doi: 10.1109/TrustCom56396.2022.00104.

[22] S. A. Khan, A. Artusi, and H. Dai, "Adversariallyrobustdeepfakemediadetectionusingfused Convolutionalneuralnetworkpredictions." [Online]. Available: https://www.kaggle.com/c/deepfake-

[23] *"2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)"*. IEEE, 2022.

[24] IEEE Signal Processing Society, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing : proceedings : April 15-20, 2018, Calgary Telus Convention Center, Calgary, Alberta, Canada*.

[25] R. Durall, M. Keuper, F.-J. Pfreundt, and J. Keuper, "Unmasking DeepFakes with simple Features," Nov. 2019, [Online]. Available: http://arxiv.org/abs/1911.00686

[26] B. Wang, X. Wu, Y. Tang, Y. Ma, Z. Shan, and F. Wei, "Frequency Domain Filtered Residual Network for Deepfake Detection," *Mathematics*, vol. 11, no. 4, pp. 1–13, 2023, doi: 10.3390/math11040816.

[27] P. Yu, J. Fei, Z. Xia, Z. Zhou, and J. Weng, "Improving Generalization by Commonality Learning in Face Forgery Detection," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 547–558, 2022, doi: 10.1109/TIFS.2022.3146781.

[28] S. Fung, X. Lu, C. Zhang, and C. T. Li, "DeepfakeUCL: Deepfake Detection via Unsupervised Contrastive Learning," in *Proceedings of the International Joint Conference on Neural Networks*, Institute of Electrical and Electronics Engineers Inc., Jul. 2021. doi: 10.1109/IJCNN52387.2021.9534089.

[29] L. Zhang, T. Qiao, M. Xu, N. Zheng, and S. Xie, "Unsupervised Learning-Based Framework for Deepfake Video Detection," *IEEE Trans. Multimed.*, 2022, doi: 10.1109/TMM.2022.3182509.

[30] B. Chen and S. Tan, "FeatureTransfer: Unsupervised Domain Adaptation for Cross-Domain Deepfake Detection," *Secur. Commun. Networks*, vol. 2021, 2021, doi: 10.1155/2021/9942754.

[31] M. S. Rana and A. H. Sung, "DeepfakeStack: A Deep Ensemble-based Learning Technique for Deepfake Detection," in *Proceedings - 2020 7th IEEE International Conference on Cyber Security and Cloud Computing and 2020 6th IEEE International Conference on Edge Computing and Scalable Cloud, CSCloud-EdgeCom 2020*, Institute of Electrical and Electronics Engineers Inc., Aug. 2020, pp. 70–75. doi: 10.1109/CSCloud-EdgeCom49738.2020.00021.

[32] Institute of Electrical and Electronics Engineers and IEEE Signal Processing Society, *2019 IEEE International Conference on Image Processing (ICIP) : proceedings : September 22-25, 2019, Taipei International Convention Center (TICC), Taipei, Taiwan*.

Authorized licensed use limited to: Manipal University. Downloaded on September 24,2024 at 09:18:45 UTC from IEEE Xplore. Restrictions apply.