International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

FACE FORENSICS – LEARNING TO DETECT MANIPULATED HUMAN FACES IN VIDEOS

ADARSH KUMAR RAY & MOHAMMED SOHAIB AHMER

Department of Computer Science & Engineering, JBIET.

Abstract – Deepfakes are artificial intelligence-synthesized content that can also fall into two other categories, i.e., lip-sync and puppet-master. Lip-sync deepfakes refer to videos that are modified to make the mouth movements consistent with an audio recording. Puppet-master deepfakes include videos of a target person (puppet) who is animated following the facial expressions, eye and head movements of another person (master) sitting in front of a camera.

1.INTRODUCTION

As the democratization of creating realistic digital humans has positive implications, there is also positive use of deepfakes such as their applications in visual effects, digital avatars, snapchat filters, creating voices of those who have lost theirs or updating episodes of movies without reshooting them. However, the number of malicious uses of deepfakes largely dominates that of the positive ones. The development of advanced deep neural networks and the availability of large amount of data have made the forged images and videos almost indistinguishable to humans and even to sophisticated computer algorithms. The process of creating those manipulated images and videos is also much simpler today as it needs as little as an identity photo or a short video of a target individual. Less and less effort is required to produce a stunningly convincing tempered footage. Recent advances can even create a deepfake with just a still image. Deepfakes therefore can be a threat affecting not only public figures but also ordinary people.

2. Body of Paper

Deepfake videos are manipulated video clips which were first created by a Reddit user, deepfake, who used TensorFlow, image search engines, social media websites and public video footage to insert someone else's face onto pre-existing videos frame by frame. Although some benign deepfake videos exist, they remain a minority. So far, the released tools that generate deepfake videos have been broadly used to create fake celebrity pornographic videos or revenge porn. This kind of pornography has already been banned by sites including Reddit, Twitter, and Pornhub. The realistic nature of deepfake videos also makes them a target for generation of pedopornographic material, fake news, fake surveillance videos, and malicious hoaxes. These fake videos have already been used to create political tensions and they are being taken into account by governmental entities. As presented in the Malicious AI report, researchers in artificial intelligence should always reflect on the dual use nature of their work, allowing misuse considerations to influence research priorities and norms. Given the severity of the malicious attack vectors that deepfakes have caused, in this paper we present a novel solution for the detection of this kind of video.

A deep learning method to detect deepfakes based on the artifacts observed during the face warping step of the deepfake generation algorithms was proposed in. The proposed method is evaluated on two deepfake data sets, namely the UADFV and DeepfakeTIMIT. The UADFV data set contains 49 real videos and 49 fake videos with 32,752 frames in total. The DeepfakeTIMIT data set includes a set of low quality videos of 64 x 64 size and another set of high quality videos of 128 x 128 with totally 10,537 pristine images and 34,023 fabricated images extracted from 320 videos for each quality set. Performance of the proposed method is compared with other prevalent methods such as two deepfake detection MesoNet methods, i.e. Meso4 and MesoInception-4, HeadPose, and the face tampering detection method two-stream NN. Advantage of the proposed method is that it needs not to generate deepfake videos as negative examples before training the detection models. Instead, the negative examples are generated dynamically by extracting the face region of the original image and aligning it into multiple scales before applying Gaussian blur to a scaled image of random pick and warping back to the original image. This reduces a large amount of time and computational resources compared to other methods, which require deepfakes are generated in advance. Recently, Nguyen et al proposed the use of capsule networks for detecting manipulated images and videos. The capsule network was initially proposed to address limitations of CNNs when applied to inverse graphics tasks, which aim

to find physical processes used to produce images of the world

3. CONCLUSIONS

Being able to detect whether a video contains manipulated content is nowadays of paramount importance, given the significant impact of videos in everyday life and in mass communications. In this vein, we tackle the detection of facial manipulation in video sequences, targeting classical computer graphics as well as deep learning generated fake videos. The proposed method takes inspiration from the family of EfficientNet models and improves upon a recently proposed solution, investigating an ensemble of models trained using two main concepts: (i) an attention mechanism which generates a human comprehensible inference of the model, increasing the learning capability of the network at the same time; (ii) a triplet Siamese training strategy which extracts deep features from data to achieve better classification performances. Results evaluated over two publicly available datasets containing almost 120 000 videos reveals the proposed ensemble strategy as a valid solution for the goal of facial manipulation detection. Using a purely AI-based approach in deepfake detection has brought the expected results. The algorithm correctly classifies videos, is lightweight and can be implemented into more sophisticated forgery tools. After the end of the competition it turned out that none of the top performing solutions used standard or biologically-inspired approaches. This suggest that AI generated content may be battled only by AI solutions, which was the idea behind choosing the XceptionNet.

ACKNOWLEDGEMENT

At outset we express our gratitude to almighty lord for showering his grace and blessings upon us to complete this Main Project. Although our name appears on the cover of this book, many people had contributed in some form or the other to this project Development. We could not have done this Project without the assistance or support of each of the following. First of all we are highly indebted to Dr. P. C. KRISHNAMACHARY, Principal for giving us the permission to carry out this Main Project. We would like to thank Dr. P. SRINIVASA RAO, Professor & Head of the Department of COMPUTER SCIENCE AND ENGINEERING, for being moral support throughout the period of the study in the Department. We are grateful to Mrs. Shaik Asha, Assistant Professor COMPUTER SCIENCE ENGINEERING, for her valuable suggestions and guidance given by her during the execution of this Project work. We would like to thank Teaching and Non-Teaching Staff of Department of Computer Science & Engineering for sharing their knowledge with us.

REFERENCES

[1] M. Zollhfer, J. Thies, P. Garrido, D. Bradley, T. Beeler, P. Prez, M. Stamminger, M. Niener, and C. Theobalt, “State of the art on monocular 3d face reconstruction, tracking, and applications,” Computer Graphics Forum, vol. 37, pp. 523–550, 2018. [2] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, “Face2face: Real-time face capture and reenactment of rgb videos,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2387–2395. [3] J. Thies, M. Zollhofer, and M. Nießner, “Deferred neural rendering: ¨ Image synthesis using neural textures,” ACM Transactions on Graphics (TOG), vol. 38, no. 4, pp. 1–12, 2019. [4] “Deepfakes github,” https://github.com/deepfakes/faceswap. [5] “Faceswap,” https://github.com/MarekKowalski/FaceSwap/. [6] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, “Protecting world leaders against deep fakes,” in IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019. [7] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, “Vision of the unseen: Current trends and challenges in digital image and video forensics,” ACM Computing Surveys, vol. 43, no. 26, pp. 1–42, 2011. [8] S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, “An overview on video forensics,” APSIPA Transactions on Signal and Information Processing, vol. 1, p. e2, 2012. [9] M. C. Stamm, Min Wu, and K. J. R. Liu, “Information forensics: An overview of the first decade,” IEEE Access, vol. 1, pp. 167–200, 2013. [10] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, “Codec and gop identification in double compressed videos,” IEEE Transactions on Image Processing (TIP), vol. 25, pp. 2298– 2310, 2016.