From puppet-master creation to false detection

Ana Pantelić and Ana Gavrovska, *Member, IEEE*

***Abstract*— Nowadays, many issues in society are affected by the misuse of deepfakes. One can say that we came to a point when prior knowledge of image processing is not a requirement for deepfake creation. With different motives, in a short period, and with limited resources, many deepfakes can appear on the internet. That brings us to testing that hypothesis of how easy and how fast someone can make a deepfake. In this paper several puppet-master creations are made for experimental purposes. In the aftermath of deepfake creation, an off-shelf available deepfake detection algorithm is applied for the detection analysis which is expected not to be universal solution for every type of deepfake realization. This brings us to high false detection, where specific cases are considered in this paper, like closed eye and head shape effects.**

***Index Terms*— Deepfake, puppet-master, deep learning, closed eye, head shape, false detection.**

I. INTRODUCTION

Technologies are rapidly evolving, and the challenges due to hardware limitation are becoming obsoleted. On the other side, different associations are trying to solve technical issues through hackathons with solutions in the interest of society. One of those issues is the detection of deepfakes [1]-[3].

DeepFakes are getting easier to produce and harder to detect. DeepFake can be considered as altering approach based on artificial intelligence and deep learning architecture. Moreover, there are different types of deepfakes, where deepfake can be often described as a synthetic switch of identities of two persons, for example in a video sequence. Namely, there are different types of deepfakes like: face-swap, entire face synthesis, puppet-mastery, and lip-syncing [4]. It is highly used in revenge pornography, and based on DEEPTRACE research [5]-[6], in September 2019, 96% of deepfake videos belong to pornographic content, where the victims are widely women. There are also widely used for politicians and public figures. With available software tool almost everyone can generate a deepfake or deepface using recorded video and an image of a taget person.

In this paper a puppet-master creation as a popular method for deepfake creation is applied. Here, performed steps for creating a deepfake is explained. Moreover, one of the methods for deepfake detection is implemented in order to observe false detections. One may have in mind that the algorithm taken for experimental analysis is not selected purposely, but in a random manner from available recent solutions, in order to observe expected false detections. It is to expect that the detection method is not prepared for dealing with each type of deepfakes and scenarios. Thus, the motivation of this work is to perform popular deepfake generation and observe what would happen if a deepfake creation approach is not directly connected to some of the state-of-the art solutions focusing on specific details like edges around important face parts like: mouth, eye and similar.

The paper is organized as follows. After introduction, in Section II we give a brief description of popular deepfake creation and detection. Section III is dedicated to simulations for puppet-master creation and for neural network based detection without taking into account the type of creation. This is followed by the experimental results and discussion related to observed detection results in Section IV. Final conclusions are given in Section V.

Ana Pantelić is with the University of Belgrade - School of Electrical Engineering, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia (e-mail: anaapantelic@gmail.com).

Ana Gavrovska is with the University of Belgrade - School of Electrical Engineering, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia (e-mails: anaga777@gmail.com; anaga777@etf.rs).

II. DEEPFAKE CREATION AND DETECTION

One of the most popular ways of creating deepfake is GAN (Generative Adversarial Network) [7]-[8]. GAN is an algorithm with two opposed neural networks that generate new, synthetic data that can pass as regular data. Neural networks and, generally, machine learning tools show the ability to mimic the human brain by learning, memorising and making the data they acquire in general. Typical deepfake algorithms for generating data are X2Face and First order motion model [2]-[3], [9].

In this paper, the neural network starts with the Monkey-Net neural network [10], and advance it to the First order motion model for image animation [9]. The First order motion model presents a fast and effective way of creating a deepfake with better results than advanced animation processing software.

Puppet-master deepfake creation is one of the modest and popular methods for making a deepfake. It shows how realistic the results can be, where the artefacts are still seen by human eye especially in video material. So, this method for creating a deepfake has its positive and negative sides. Artefacts are observable and this can be negative experience for the creater. This is also a positive information since we can still believe that we can distinguish true or false video story.

For deepfake detection, one has to be aware of the algorithm used for deepfake creation. Taking into consideration that the person who published a deepfake won't leave a piece of information related to origin or source or used tools for deepfake design, the common decision is that the detection from practial point of view will be applied to images cropped/grabbed from the video.

The main focus of detection of a deepfake are face parts like eyes and mouth, and head movement. Deepfake may be recognized on irregular pixel weight at the edges of the mentioned regions after training the neural network that detects them. One of the state-of-the art models based on such detection is Meso-4 model [11]-[12]. It is based on convolutional neural network and represents an efficient tool for dealing with particular types of visual modifications.

 ISBN 978-86-7466-930-3

Here, it is used as only one of the solutions for observing how can some deepfake examples created by the First order motion model be detectable as deepfakes and what are the cases when one may get high number of false detections, when deepfake frames are passed as regular images.

III. SIMULATION

A. *Steps for creating a deepfake*

Here, preparations for the deepfake creating is filming a video of 256×256 px, that is frontal with simple backround. Next to that, an image of a person we want to switch the identity with is chosen. The selected image is also frontal with simple backround and with the lack of face-covering details. In this paper Google Colaboratory with python script are used for experimental purposes [13]-[14].

First order motion model for image animation is upgraded convolutional neural network based on Monkey-Net-u [9], [11]. Monkey-Net codes information of movement through keypoints that are pretrained and self-observed. Down side of this neural network is possibly that, while following the trajectory of keypoints in faster movements, missing spots in image can appear. To solve that missing spot, affine transformation is used to perserve the propotion of the line and collinearity between dots. While it perserves propositions between lines and dots, affine transformation doesn't perserves all the angles. To have a clear overview, the improved parts of Monkey-Net are: keypoints aren't just detected but they are self-controled, which will add the adaptability on the targeted image; the generator of occlusion is added that gives a mask based on the parts that are not initially visible; it improved the visual domain of the puppet in general. In Fig. 1 simulation steps for creating a puppet deepfake are shown. If we follow from the video of the puppet, the keypoints are generated and local affine transformation is applied and collected. Until it comes to dense motion, from the source image we encode all the needed features and by warp operation the parts of process are connected. This is followed by occlusion map and decoder that brings the final creation result.

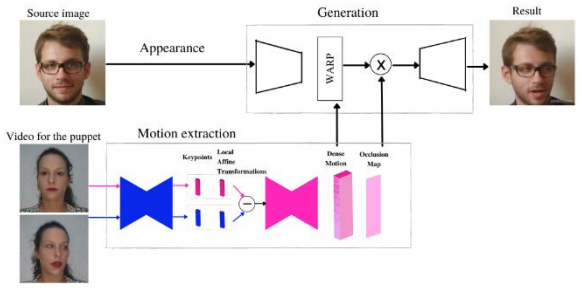

Fig. 1. Simulation steps for creating a puppet deepfake.

B. *A deepfake detection model*

For the need of deepfake detection, a pretrained convolutional neural metwork MESO-4 model is used [12]. The architecture of the model is shown in Fig. 2. It is constructed of four convolutional blocks and one hidden layer. It recognizes the vertical and horizontal lines; applies batch normalization; uses the convolutional matrix with a task to bring all the important edges like edge of the lips, shape of the head or edges of the eyes; it employs pooling layer that will pick the pixel with dominating characteristics. Based on the pixel the reducing of spatial size of convolutional operations is possible. To improve the generalization, in the addition to the normalization of the batch, ReLU (Rectified Linear Unit) which introduces nonlinearity is applied.

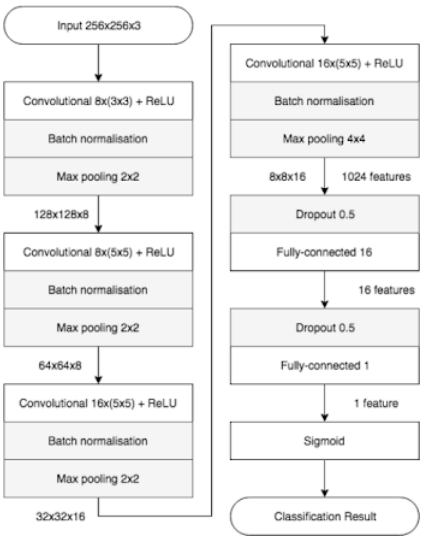

Fig. 2. Meso-4 architecture, where layers and parameters are displayed in rectangles [12].

C. *False detection counting in a puppet related data*

We wanted to see if the head movement and misshapes of the head will be detected by Meso-4 neural network. That is the reason we produced the video with head movement on both sides (left and right), and then abruptly moved to one of the sides. Furthermore, the idea is to make mouth and eyes to seem natural in a deepfake, so the person often blinks. When making mouth movements, the mouth is moved naturally without exaggeration.

IV. EXPERIMENTAL RESULTS

A. *Experimental results for created deepfakes and false detection*

Three deepfake videos are made with source images of university assistant, public figure, and politician. One of the created deepfakes for a public figure is shown in Fig. 3.

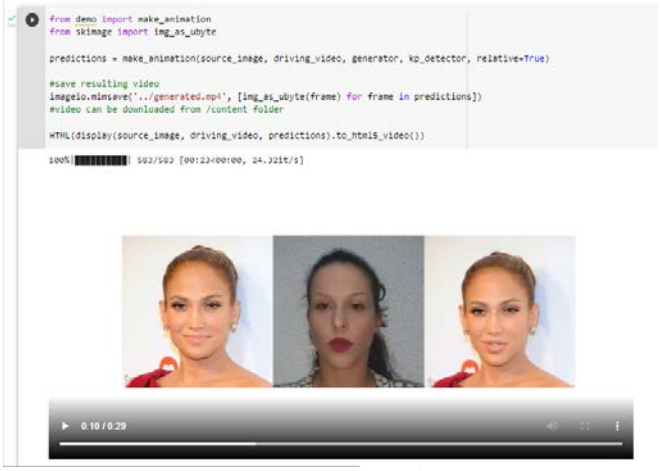

Fig. 3. One of the created deepfakes. From left we have source image, video for the puppet and the end result.

 ISBN 978-86-7466-930-3

The results were satisfying, with expected characteristics of puppet-mastery. The process resulted in a source image that was following all the given movements of head, eyes, and lips.

For the detection task both original and deepfake frames are used. Totally, 461 images are tested whether they are real or not, giving for original examples satisfying results as shown in Fig. 4. Predicted likelihood is close to 1 which means that labeling is performed in adequate manner. Correct prediction can be noticed and it proves that the result is true.

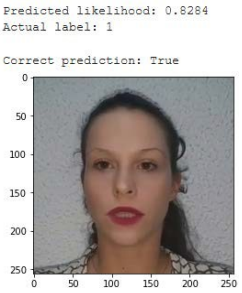

Fig. 4. The result of the detection. If the predicition is closer to one it means it sees it as real, and next to that it proves it predicition as True or False.

At first glance, a lot of images that are real are also detected as real, but unsatisfactory results are shown for deepfakes. The summary of obtained results is given in Fig.5. Out of 110 real images, 94 real cases are detected. On the other hand, high percentage is found for false detections where deepfake frame is considered real. In the experiment it is found that we have around 82.33% of chances for misinterpreting a deepfake image as real, and around of 14.5% of misinterpreting a real as a deepfake. The high false detection exists as it was expected.

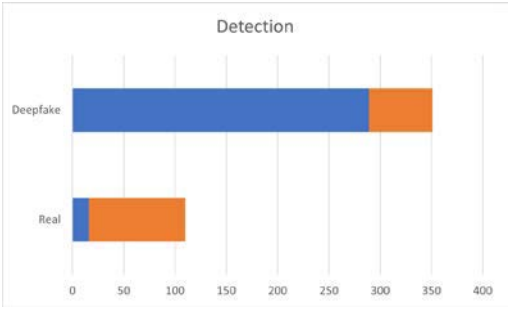

Fig. 5. The summary of detection. The blue colour presents the number of falsely detected images. Out of 351 deepfake images, 289 are detected real.

B. Further analysis on accurate results

When observing accurate real results it was noticed that the probability of images with closed eyes, and when the head is curved, is close to 0.5, which means that there is a significant level of doubt. Similarly is noticed with cases where errors occurred, i.e. when real images are detected as deepfakes. The probability of around 50% can be interpreted as random class selection.

In Fig. 6 examples of true predictions are presented, where one of the examples show lower predicted likelihood. This is the case where eyelids and pupils are not visible while blinking. On the other hand, when eyelids and pupils are visible and when there is less blur around the face, there is a higher probability that the detected image is real.

Fig. 6. Examples of True predicitions.

There are frames that are detected as deepfake and similar pattern can be recognized as in Fig. 7.

Fig. 7. Real images detected as deepfake

It can be noted that there are frames in Fig. 7. where pupils and irises are not visible, as well as where there is a greater curvature of the head. In these cases Meso-4 has detected that there is a chance that it is a deepfake. Also, it is important to note that most predictions are around 0.4 and that the probability leans towards deepfake, mostly where the eyes are closed and the head is tilted. The results of the experiment were best shown on the targeted personalities with similar facial symmetry as in the video used for the puppet.

The script for the puppet video had slow head movements to one side and then to the other, and then it abruptly moved to one side to see the artefact mentioned in the works for Monkey-Net and the First Order Motion Model. The appearance of this artefact was expected, and the artefact appeared in the results. For a better solution, photos from other angles should be found for the target person, which would improve the occlusion - in the sense that the focus is placed on the other eye (which should not be visible) or the whole part of the face that should be covered, and a 3D model, which is not obtained here due to the different shape of the person's head in the original. As for the background, the improved occlusion gave favourable results. Also, on several clippings, an artefact was obtained, which was conceptually mitigated, and that is the disappearance of parts of the image/person, as shown in Fig.8.

Fig. 8. Image with the disappeared part of the image

 ISBN 978-86-7466-930-3

Accurate results are obtained in this case, and a significant number of cases, as shown in Fig.9. It can be immediately noticed that the third of the displays are around 0.4 probabilities. All detected images have distorted heads and artefacts due to movement, i.e. tracking the trajectory of key points on the face of different symmetry in the 3D model or they have their eyes closed.

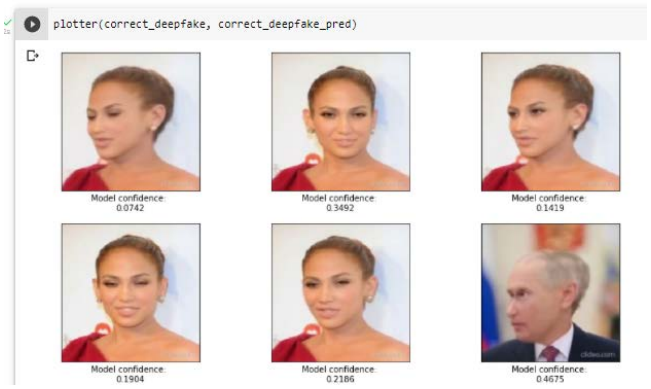

Fig. 9. Examples of images that are deepfakes, and detected as deepfakes

C. *Analysis of false detection results*

Here, of the greatest importance are images that are detected as real, and these are images from deepfake videos.

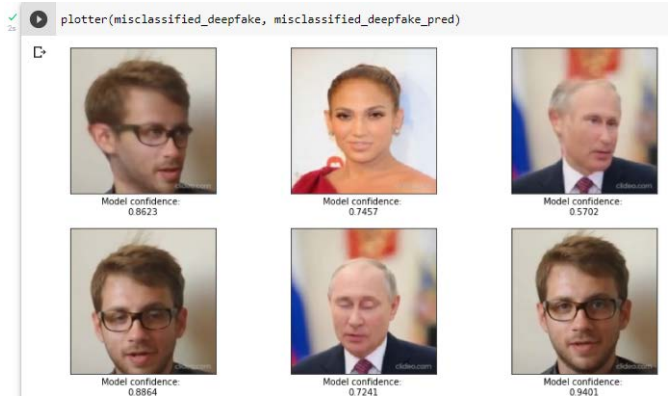

Fig. 10. Examples of images that are deepfakes and detected as real.

Three specific cases can be observed where there is a space for further improvements:
- mouth/lip movements,
- head movements and
- eye movements

Speaking of mouth movements, Fig.11, satisfactory results were obtained in creating a puppet, where the targeted persons followed the movements and had a tooth display at the appropriate moments. All lip movements that were observed in isolation from other movements contributed to the reality of deepfake and false detection. The lips are mostly in motion with teeth or collected. Compared to well-detected lips, every fifth well-detected mouth as a deepfake has a deformity along with the head. So one can say that lips are one of the parts where there is a need for a better detection.

Slow head movements to one side and then the other, and then abruptly movement to one side is a significant process, Fig.12. All the end movements are recognized as real images, in a manner that if the head is still, meaning in one place for more than one second without motion blur, the image will be detected as real. Furthermore, the detector sees even distinguished distortions of the head and face as realistic images. This is noticed in nearly 33% of false detected images. Moreover, it is very likely to have errors in differentiation when real images are frontal with clear head movement, while eyes are visible and lips in motion or collected. Those characteristics are visible in around 97% of all images that are deepfake and marked as real.

Fig. 11. Images with mouth movement

Fig. 12. Examples of frontal, right, left and right side (column-wise).

The most relevant results in terms of authenticity are in creating obtained by tracking eye movements, Fig.13. It can be considered the most difficult case in the creation process. False detection occurred in the eye related situations when the eyes are fully visible or when the iris and pupil are visible. Also, false detection is present when the eyes have a proportional distance between the pupils concerning the position of the face. This is shown in the most majority of images/frames, around 99%.

Fig. 12. Examples of eye movement.

 ISBN 978-86-7466-930-3

V. CONCLUSION

Based on the results of the experiments, it can be concluded that the creation of fast and efficient deepfake still meets the need for additional training of neural networks that map key points and where trajectories needs to be adapted to different head shapes so that the appearance of deepfake is invisible to the human eye. It is important to emphasize that the results completely coincided with the results from the paper [9]. It is believed that further training of such neural networks can lead to even more adaptable results.

On the other hand, the selected detection was not the best example for a given set of images which is shown through a variety of false detection results. This can be attributed to the fact that the detection follows certain edges and looks for dominant characteristics that were not the focus in creating the deepfake, such as eye details. Therefore, it is suggested that the type of detection should be adapted to specific type of deepfake. The importance of focusing on deepfake detection, and particularly eye movement, is emphasized which must be adapted to all new ways of creating deepfake videos.

In future work, we would focus on better detection of the misshaped head, as well as different tools for recognition of lip-syncing and eye tracking.

ACKNOWLEDGMENT

This work is written during the research supported and partially funded by the Ministry of Education, Science and Technological Development, Republic of Serbia. No. 2022/200103.

REFERENCES

[1] A.M. Almars, "Deepfakes detection techniques using deep learning: a survey," *Journal of Computer and Communications*, vol. 9, no. 5, pp. 20-35, May 2021.

[2] J.T. Hancock, and J.N. Bailenson, "The social impact of deepfakes," *Cyberpsychology, behavior, and social networking*, vol. 24, no. 3, pp. 149-152, 2021.

[3] M. Đorđević, M. Milivojević, and A. Gavrovska, "DeepFake video production and SIFT-based analysis," *Telfor Journal*, vol. 12, no. 1, pp. 22-27, 2020.

[4] 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," *arXiv preprint arXiv:2103.00484*, 2021.

[5] C. Gosse, and J. Burkell, "Politics and porn: how news media characterizes problems presented by deepfakes," *Critical Studies in Media Communication*, vol. 37, no. 5, pp. 497-511, 2020.

[6] J.P. Dasilva, K.M. Ayerdi, and T.M. Galdospin, "Deepfakes on Twitter: Which Actors Control Their Spread?," *Media and Communication*, vol. 9, no. 1, p. 301-312, 2021.

[7] L. Guarnera, O. Giudice, and S. Battiato, "Deepfake detection by analyzing convolutional traces," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 14-19 June, 2020. Doi: 10.1109/CVPRW50498.2020.00341

[8] C. Yang, L. Ding, Y. Chen, and H. Li, "Defending against gan-based deepfake attacks via transformation-aware adversarial faces," In *2021 International Joint Conference on Neural Networks* (IJCNN) IEEE, pp. 1-8, July 2021.

[9] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "First order motion model for image animation," *Advances in Neural Information Processing Systems* 32, 2019.

[10] M.T. Jafar, M. Ababneh, M. Al-Zoube, and A. Elhassan, "Forensics and analysis of deepfake videos," In *2020 11th international conference on information and communication systems* (ICICS) IEEE, pp. 053-058, April 2020.

[11] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe. "Animating arbitrary objects via deep motion transfer," In *CVPR*, pp. 2377-2386, 2019.

[12] 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. https://doi.org/10.1109/WIFS.2018.8630761

[13] Google Colab, https://colab.research.google.com/

[14] Python, https://www.python.org/

 ISBN 978-86-7466-930-3