Journal of Computing and Communication Vol.2 , No.2 , PP. 31-37 , 2023

DeepFakeDG: A Deep Learning Approach for Deep Fake Detection and Generation

Zeina Ayman [a], Natalie Sherif [a], Mariam Mohamed [a], Mohamed Hazem [a], Diaa Salama *[a,b]

[a] *Department of Computer Science, Faculty of Computer Science, Misr International University, Cairo, Egypt*

[b] *Department of Information Systems, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt*

*Corresponding Author: Diaa Salama [***diaa.salama@miuegypt.edu.eg***]

ARTICLE DATA

Article history:
Received 27 May 2023
Revised 25 June 2023
Accepted 01 July 2023
Available online

Keywords:
Neural Machine Translation
Sequence to Sequence Model
Sign Language
Deep learning
Transformer

ABSTRACT

The main idea of this project is to develop a web application that can help to detect whether the input data we provide of people, whether celebrities or people in general, are real or fake and generate deepfakes themselves. Recently, with the evolution of technology and advanced image editing tools, people can easily get manipulated, as deepfake algorithms can easily create fake videos and images that people can't distinguish from authentic ones, an emerging problem threatening the trustworthiness of online information. Deepfakes mainly affect public figures, celebrities, and politicians. Forged videos are videos that contain fake images over real ones. In this research, there are methods used with Machine and deep learning approaches that will be used with the dataset composed of deep fake videos and authentic ones to detect these manipulations and protect the government from criminals. There will be various techniques used to distinguish real from fake using face swapping, or is there something off regarding its behavior, or if a voice of a person is used with another person's voice, etc. The deep fake detector can be used in courts and police stations to reduce the likelihood of crimes and frauds that may happen and detect them. This project aims to make a website to detect whether videos are fake or not. More and above, the proposed model will also provide a deepfake generation efficiency.

1. Introduction

Deepfake is a technique that replaces the face of a targeted person in a video with the face of someone else by joining a synthesized face region into the original image. It appeared to publicity in 2017 [1] when a Reddit user posted videos of celebrities in sexual positions; also, deepfake was used on President Barack Obama, accusing him of stating things he didn't say. Later, several applications emerged that produced deepfake content, and they were created to interfere with public opinion and political campaigns. Deepfakes target social media platforms where rumors can be spread easily as users tend to follow the flow. Deepfake has an increasingly harmful effect on personal privacy and social security; it can damage the reputations of celebrities and public figures and spread rumors worldwide [2]. Not only famous people could be affected, but also it can be anyone around so that the person who created fake content of theirs could blackmail them; deep fakes involved a large number of famous political leaders, actresses, comedians, and entertainers who had their faces stolen and used in unreal videos causing them huge problems. Adding to that, this system can generate a deepfake. [3]

2. Background

Deepfake is a technique that replaces the face of a targeted person in a video with the face of someone else by joining the synthesized face region into the original image. It appeared to publicity in 2017 when a

Reddit user posted videos of celebrities in sexual positions; also, deepfake was used on President Barack Obama, accusing him of stating things he didn't say. Later, several applications emerged that produced deepfake content, and they were created to interfere with public opinion and political campaigns. Deepfakes target social media platforms where rumors can spread easily as users follow the flow. Deepfake has an increasingly harmful effect on personal privacy and social security; it can damage the reputations of celebrities and public figures and spread rumors worldwide. Not only famous people could be affected, but also it can be anyone around so that the person who created fake content of theirs could blackmail them; deep fakes involved a large number of famous political leaders, actresses, comedians, and entertainers who had their faces stolen and used in unreal videos, causing them huge problems.

3. Related Work

3.1. Deressa Wodajo, Solomon Atnafu[2] :

Agreed that deep fakes cause a significant threat to everyone if used for harmful purposes such as phishing, scam, and identity theft, reducing the trustworthiness of the public data. A convolutional vision transformer is used in this project to detect the deep fake. This project adds a CNN module to the VIT architecture as CNN extracts the features (facial features in an image), and VIT takes those features as input to categorize them into a specific class. the datasets used are FaceForensics++Faceswap,FaceForensics++deepfakedetaction,FaceForensics++deepfake,FaceForensics+ faceshifter, FaceForensics++ neuraltextures. the results are An accuracy of 91.5%, an AUC value of 0.91, and a loss value of 0.32, which indicates the difference between the predictions and the actual results. CNN and RNN had an accuracy of 92.6% (validation) and 91.88% (testing). CViT had 87.25 (validation) and 91.5(testing). The face recognition library had the best results compared to MTCNN and BlazeFace libraries as it had higher accuracy.

3.2. Yuezun Li, Xin Yang, Pu Sun [4] :

Discussed about AI-synthesized face-swapping videos, commonly known as DeepFakes, are an emerging problem threatening the trustworthiness of online information. Developing and evaluating DeepFake detection algorithms calls for large-scale datasets is needed. The dataset used is the Celeb-DF dataset. Celeb-DF is generally the most challenging to the current detection methods, and their overall performance on Celeb-DF is lowest across all datasets, with an average AUC of 56.9 %. At the same time, FF-DF had the highest results across all methods, with an average AUC of 82.3 %. While the method that had the highest performance across all datasets was DSP-FWA which had an AUC of 87.4 %, and the lowest performance goes to the HeadPose method, with an AUC of 58.7 %. The celeb-DF dataset has proven that it still needs improvement.

3.3. Darius Afchar, Vincent Nozick, Junichi Yamagishi, Isao Echizen [5] :

Agreed that the huge use of digital images has been followed by a rise of methods to change image contents, using editing software like Photoshop. The digital image forensics research field is dedicated to detecting fake images to regulate the circulation of such fake content. Providing two possible network architectures(Meso 4 and MesoInception 4) to detect such forgeries efficiently with a low computational cost. The dataset used is the Deepfake dataset, the Face2Face dataset. Both networks have reached close scores, around 90 %, considering each frame independently. A higher score is not expected as some images have facial extractions with a very low resolution. It is observed a decline of scores at the strong video compression level. The image aggregation significantly enhanced both detection rates. It even rose greater

than 98 % with the MesoInception, network on the Deepfake dataset. Note that on the Face2Face dataset, the score is the same for both networks, but the misclassified videos are dissimilar.

3.4. Xin Yang, Yuezun Li, Siwei Lyu [6] :

Agreed that Deep faking greatly impacts our environment nowadays; it's created by putting faces using deep neural networks into original images/videos. Together with additional forms of misinformation shared through the digital social network, digital impersonations were created by deepfake, which have become a real problem with negative social influence. Accordingly, there is a serious need for successful methods to detect Deep Fakes. The dataset used is UADFV and a subset of the DARPA. The results are the SVM classifier reaches an AUROC of 0.89. This means that the difference between head poses evaluated from the central region and the whole face is a good feature for identifying Deep Fake generated photos. DARPA GAN Challenge dataset, the AUROC of the SVM classifier is 0.843. The reason is that the synthesized faces in the DARPA GAN challenges are mostly blurry, leading to a struggle to predict facial landmark places accurately and, as a result, the head pose evaluation. They also calculated the performance using separate videos as the unit of analysis for the UADFV dataset. This is achieved by taking the average of the classification prediction on frames over separate videos. Also, other related work in [7-29] has been proposed in recent years to address machine learning and its application in different fields.

4. Proposed Approach

This section demonstrates the proposed method for detecting and generating Deep Fakes.

4.1. System Overview

As shown in Figure 1, the following describes the system overview. The system is separated into two parts detection and generation. As for the detection side, the dataset is used, and pre-processing is done to extract faces from the videos/images, which are then passed to a machine learning classifier (cnn, vgg,etc.), and training is done on the imported dataset. The user uploads a video/image to the website to detect it, and the result is shown on the user interface. As for the generation side, the user uploads two videos, and the model does frames extraction on both videos (the source and the target), then facets are extracted from these frames. They are gathered to be trained to match the face of the source video to the face of the target video. The face is merged with the destination video, and several enhancements can be done as blur, sharpening, color correction, etc. Finally, the deepfake video is generated.

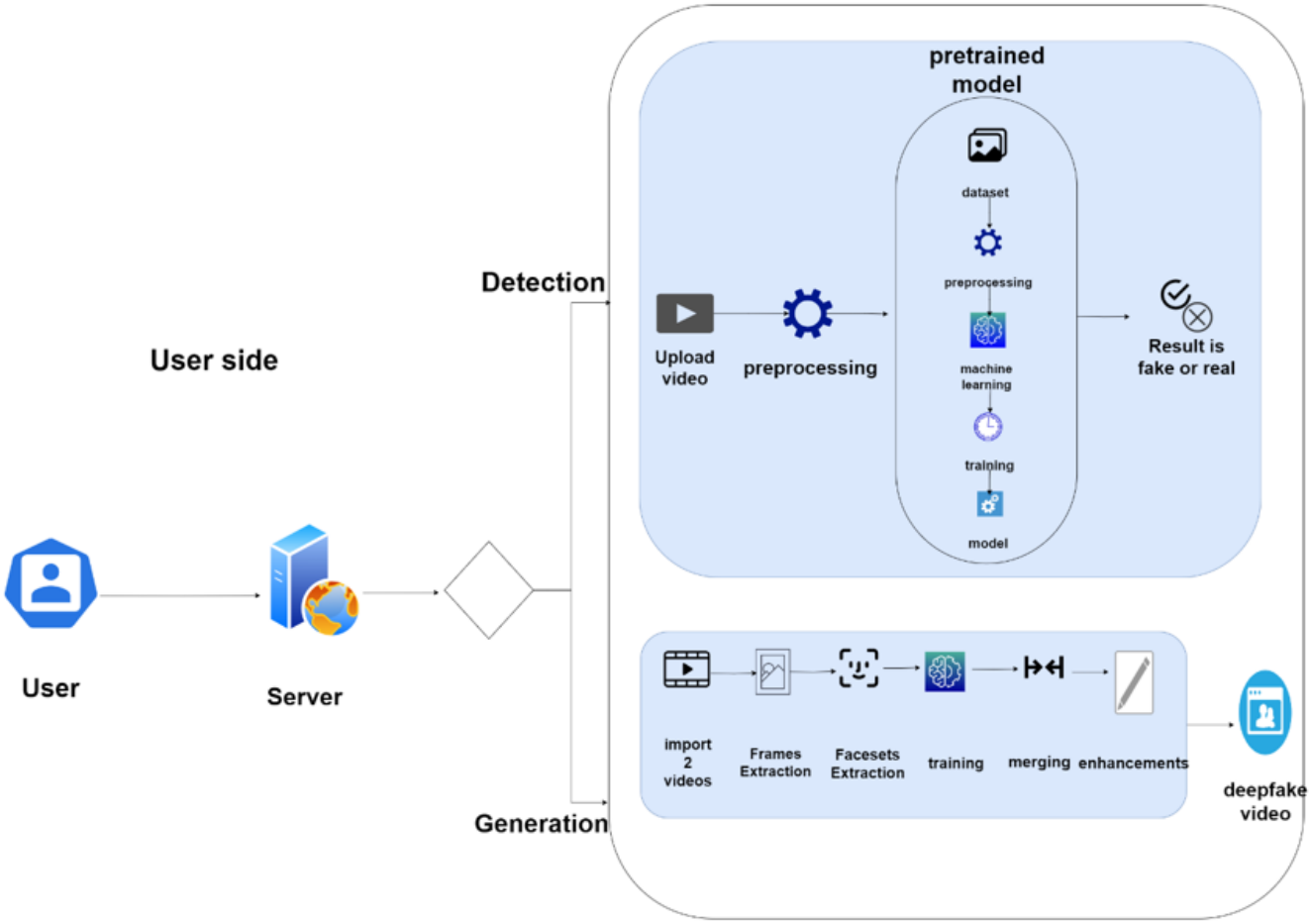

Figure 1 Deepfake detection and generation

5. Methodology

5.1. CNN:

A convolutional neural network (CNN or ConvNet is a network architecture for deep learning that learns directly from the data below in Fig. CNNs are useful for finding image patterns to recognize objects, classes, and categories.[7] Convolutional Neural Networks comprise node layers, convolution, pooling, hidden, and output layers. Our image passes through all these layers, and every step is important. As for the convolution layer, this layer is the main block of CNN. This is the layer where most of the computations occur. This layer has three inputs. The first is input data, the second is a filter, and the third is a feature map. The convolution process consists of the input image used with filters, and filters are mainly 3x3 matrices that are iterated over the image. In the end, an activation function calculates the final result. The output is stored in an output matrix.[8] The pooling layer is used for downsampling, reducing the number of parameters in our input image. The fully connected layer is the layer that performs classification on the image and decides whether an image is real or fake. Using the image data generator, we used a target size of 224 x 224 and a batch size 32 to build the model. A max-pooling of size two is then used to reduce the dimensions of the input image, and a sigmoid activation function is used. The model is compiled using a binary cross entropy to perform a percentage of how much fake or real is produced.[10]

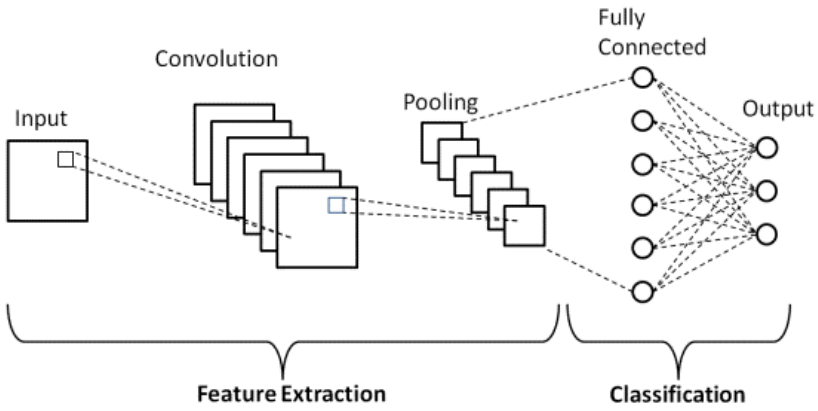

Figure 2 CNN architecture

5.2. VGG

In this paper, VGG is selected as a second architecture.[8] In the figure below it shows the flow of the VGG model. The VGG is a pre-trained neural network technique primarily used for image recognition issues. The VGG is based on the convolution neural net (CNN) architecture[9] Years ago, the VGG model architecture was first utilized in the ILSVR competition. The convolution and pooling layers of VGG can be divided into blocks, numbered Block 1 to Block 5, composed of convolution and maxpooling layers. A target size of 150 x 150 is used. The train generator used a batch size of 256, and the validation generator used a batch size of 5 [11].

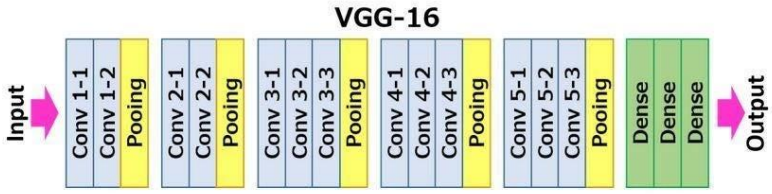

Figure 3 VGG architecture

5.3. Datasets:

The main dataset used in our paper is the Deepfake Detection Challenge (DFDC).[12] Four hundred train videos are split into 70% training and 30% validation. The dataset comprises videos of random people in different settings and lightnings, and they are recorded outdoors and indoors to generalize the dataset as much as possible. The average size of the video is 5 Mb, and the size of the training video is 4.44 Gb. They are split with fake and real labels, and a JSON file is accompanied by them, including each video's labels. A pre-processing technique is done on the videos to extract faces from videos within every frame. The images are then rescaled to 224 x 224 pixels[12].

Figure 4 Real image

Figure 5 Fake image

5.4. Experimental Results:

The following table shows the results demonstrated after using both CNN & VGG are quite similar. Still, the CNN has shown slightly better results as the accuracy was around 94% in the CNN, Unlike the VGG, which was around 88%, and after more testing for both, the loss in the VGG was higher than that in the CNN which was 27% in VGG to 16\% in CNN this shows that the results in CNN are quite better.

Table 1
Summary of the results of different algorithms

	Accuracy	**Val Accuracy**	**Loss**	**Val Accuracy**
CNN	94%	61%	16%	92%
VGG	88%	52%	27%	91%

6. Conclusion

this paper covers the detection and the generation of the deepfake, where the generation is used to help people know if the content they want is forged or not, and on the other hand, we could help them generate the deepfake itself. The detection showed its best performance with the CNN, and more accurate results are aimed to be established in the future for a more authentic, precise website. It is also intended to work with different datasets and generalize the dataset more with several augmentation techniques so that the system can detect any data inserted by the user.

References

[1] Westerlund, Mika. "The emergence of deepfake technology: A review." *Technology innovation management review* 9, no. 11 (2019).

[2] Yu, Peipeng, Zhihua Xia, Jianwei Fei, and Yujiang Lu. "A survey on deepfake video detection." *Iet Biometrics* 10, no. 6 (2021): 607-624.

[3] Maksutov, Artem A., Viacheslav O. Morozov, Aleksander A. Lavrenov, and Alexander S. Smirnov. "Methods of deepfake detection based on machine learning." In *2020 IEEE conference of russian young researchers in electrical and electronic engineering (EIConRus)*, pp. 408-411. IEEE, 2020.

[4] Li, Yuezun, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. "Celeb-df: A large-scale challenging dataset for deepfake forensics." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3207-3216. 2020.

[5] Afchar, Darius, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. "Mesonet: a compact facial video forgery detection network." In *2018 IEEE international workshop on information forensics and security (WIFS)*, pp. 1-7. IEEE, 2018.

[6] Yang, Xin, Yuezun Li, and Siwei Lyu. "Exposing deep fakes using inconsistent head poses." In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8261-8265. IEEE, 2019.

[7] Ahmed, Saadaldeen Rashid, Emrullah Sonuç, Mohammed Rashid Ahmed, and Adil Deniz Duru. "Analysis survey on deepfake detection and recognition with convolutional neural networks." In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1-7. IEEE, 2022.

[8] Nirkin, Yuval, Lior Wolf, Yosi Keller, and Tal Hassner. "Deepfake detection based on discrepancies between faces and their context." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 10 (2021): 6111-6121.

[9] Nirkin, Yuval, Lior Wolf, Yosi Keller, and Tal Hassner. "Deepfake detection based on discrepancies between faces and their context." IEEE Transactions on Pattern Analysis and Machine Intelligence 44, no. 10 (2021): 6111-6121.

[10] Zhao, Tianchen, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, and Wei Xia. "Learning self-consistency for deepfake detection." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 15023-15033. 2021.

[11] Nirkin, Yuval, Lior Wolf, Yosi Keller, and Tal Hassner. "Deepfake detection based on discrepancies between faces and their context." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 10 (2021): 6111-6121.

[12] Chang, Xu, Jian Wu, Tongfeng Yang, and Guorui Feng. "Deepfake face image detection based on improved VGG convolutional neural network." In *2020 39th chinese control conference (CCC)*, pp. 7252-7256. IEEE, 2020.

[13] Abd Elminaam, D.S., El Tanany, A., Salam, M.A. and Abd El Fattah, M., 2022, May. CPSMP_ML: Closing price Prediction of Stock Market using Machine Learning Models. *In 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 251-255)*. IEEE.

[14] Ai, Mona AS, Anitha Shanmugam, Suresh Muthusamy, Chandrasekaran Viswanathan, Hitesh Panchal, Mahendran Krishnamoorthy, Diaa Salama Abd Elminaam, and Rasha Orban. "Real-time facemask detection for preventing COVID-19 spread using transfer learning based deep neural network." *Electronics* 11, no. 14 (2022): 2250.

[15] Neggaz, N. and AbdElminaam, D.S., 2021, May. Automatic sport video mining using a novel fusion of handcrafted descriptors. *In 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 387-394)*. IEEE.

[16] Salam, M.A., Ibrahim, L. and Abdelminaam, D.S., 2021. Earthquake Prediction using Hybrid Machine Learning Techniques. *International Journal of Advanced Computer Science and Applications, 12(5), pp.654-6652021*.

[17] Mahmoud, E., Kader, H.A. and Minaam, D.A., 2019, October. Fuzzy knowledge base system for floating car data on SUMO. *In 2019 29th International Conference on Computer Theory and Applications (ICCTA) (pp. 38-42)*. IEEE.

[18] AbdElminaam, D.S., ElMasry, N., Talaat, Y., Adel, M., Hisham, A., Atef, K., Mohamed, A. and Akram, M., 2021, May. HR-chat bot: Designing and building effective interview chat-bots for fake CV detection. *In 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 403-408)*. IEEE.

[19] AbdElminaam, D.S., Fahmy, A.G., Ali, Y.M., El-Din, O.A.D. and Heidar, M., 2022, May. DeepECG: Building an Efficient Framework for Automatic Arrhythmia classification model. *In 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 203-209)*. IEEE.

[20] AbdElminaam, D.S., Fahmy, A.G., Ali, Y.M., El-Din, O.A.D., Aly, A.R. and Heidar, M., 2022, May. ESEEG: An Efficient Epileptic Seizure Detection using EEG signals based on Machine Learning Algorithms. *In 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 210-215)*. IEEE.

[21] AbdElminaam, D.S., Ahmed, N., Yasser, M., Ahmed, R., George, P. and Sahhar, M., 2022, May. DeepCorrect: Building an Efficient Framework for Auto Correction for Subjective Questions Using GRU_LSTM Deep Learning. *In 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 33-40)*. IEEE.

[22] Ali, M.A., Orban, R., Rajammal Ramasamy, R., Muthusamy, S., Subramani, S., Sekar, K., Rajeena PP, F., Gomaa, I.A.E., Abulaigh, L. and Elminaam, D.S.A., 2022. A Novel Method for Survival Prediction of Hepatocellular Carcinoma Using Feature-Selection Techniques. *Applied Sciences, 12(13), p.6427*.

[23] *AbdElminaam, D. S., El-Aal, A., & Abdellatif, A. (2023). Nowcasting Egypt GDP Using machine learning Algorithms. Journal of Computing and Communication, 2(1), 1-8.*

[24] *Radwan, M., Mohamed Abdelrahman, N., Wael Kamal, H., Khaled Abdelmonem Elewa, A., & Moataz Mohamed, A. (2023). MLHeartDisPrediction: Heart Disease Prediction using Machine Learning. Journal of Computing and Communication, 2(1), 50-65.*

[25] *Tamer Ghareeb, B., Tarek, F., & Said, H. (2023). FER_ML: Facial Emotion Recognition using Machine Learning. Journal of Computing and Communication, 2(1), 40-49.*

[26] *Essam, F., Samy, H., & Wagdy, J. (2023). MLHandwrittenRecognition: Handwritten Digit Recognition using Machine Learning Algorithms. Journal of Computing and Communication, 2(1), 9-19.*

[27] Chang, Xu, Jian Wu, Tongfeng Yang, and Guorui Feng. "Deepfake face image detection based on improved VGG convolutional neural network." In *2020 39th chinese control conference (CCC)*, pp. 7252-7256. IEEE, 2020.

[28] Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "Fighting deepfake by exposing the convolutional traces on images." *IEEE Access* 8 (2020): 165085-165098.

[29] Nirkin, Yuval, Lior Wolf, Yosi Keller, and Tal Hassner. "Deepfake detection based on discrepancies between faces and their context." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 10 (2021): 6111-6121