Detecting Deepfakes Using GAN Manipulation Defects in Human Eyes

Elisabeth Tchaptchet
Mathematics and Computer Science Department
University of Dschang
Cameroon
elitchap@gmail.com

Elie Fute Tagne
Mathematics and Computer Science Department
University of Buea
Cameroon
eliefute@yahoo.fr

Jaime Acosta
DEVCOM Army Research Laboratory
Network Security Branch
MD, USA
jaime.c.acosta.civ@army.mil

Rawat Danda
Department of Computer Science
Howard University
Washington DC, USA
danda.rawat@howard.edu

Charles Kamhoua
DEVCOM Army Research Laboratory
Network Security Branch
MD, USA
charles.a.kamhoua.civ@army.mil

***Abstract*—The Deepfake phenomenon is very important nowadays because there are possibilities to create very real images that can fool anyone, thanks to deep learning tools based on generative adversarial networks (GAN). These images are used as profile images on social media, aimed here at creating discord and scams internationally. In this work, we show that these images can be detected by a multitude of imperfections present in the synthetized eyes such as the irregular shape of the pupil and the difference between the corneal reflections of the two eyes. These imperfections are caused by the absence of physical/physiological constraints in most GAN models. We are developing a two tier architecture able of detecting these deepfake images. It starts with an automatic segmentation method of the eye pupil to check the shape. Then, for pupils of non-standard shape, the whole image is taken, transformed into gray level and then passed into an architecture that extracts and compares the corneal specular reflections of two eyes. Experimenting with a large set of real image data from the Flickr-Faces-HQ dataset and fake styleGAN2 images demonstrates the effectiveness of our method. Our method has good stability for physiological properties during deep learning; therefore, it is robust as some of the single-class deepfake detection methods. The results of the experiments on the selected datasets demonstrate greater precision compared to other methods.**

***Index Terms*—Adversarial machine learning, Deepfake, face generation, GAN**

I. Introduction

Today there is a growing phenomenon that is spreading exponentially called Deepfake: it is the possibility of generating an image from other images or automatically modifying a person's face, either by increasing the details, for example by changing the emotions, or by removing objects, details in images and videos through algorithms based on ”Deep Learning”. With this technology, it is possible to generate high-quality images to create content that cannot be easily detected by the human eye. The term ”Deepfake” refers to any content that is modified or synthetically generated using generative adversarial network models. An example is shown in Figure 1 [1], [2]. These pictures represent images of people's faces that do not exist but have been created from scratch from examples provided to GAN. The GAN is a set of neural networks with two parts: a generator model and a discriminator model, which learns from a large number of data and creates a new data resembling the ones provided at the beginning but which do not exist. These contents are used for malicious purposes by malicious users to cause serious problems in society or political threats. In order to mitigate these risks, several methods for detecting false images have been proposed [3], [4]. Most of these methods are designed with the help of deep neural networks (DNN), as they have a high precision in the field of image detection [5], [6].

Fig. 1. GAN-generated images: images from http: //thispersondoesnotexist.com generated with StyleGAN2

Because of the misuse of deepfakes contents by some malicious people, which harm the reputation of healthy people, several methods of detection have been set up

Distribution Statement A: Approved for public release. Distribution is unlimited.

to expose its contents. Several methods prove that these contents are false or modified by verifying the representation of certain parts of the face such as eyes, mouth, nose etc. These methods are called physiological/physical detection methods such as [5]–[7] and their results are generally easier to interpret. Although these methods are effective, they have two important limitations: (1) the color system in which the images are located is not the same for all images in the dataset, resulting in many false positives during detection, (2) uncorrected illumination of the images often leads to cases of over-illuminated or partially illuminated images, resulting in poor detection. Marten et al. [8] developed a method exposing deepfakes that uses the fact that in some of the generated contents there are missing elements on the face, like the reflection of light in the eyes, the tooth area is not well represented. Hu et al. [4] expose these generated contents by showing this big difference in the eyes of the deepfakes taking into account that real images usually have a resemblance in the two corneas of the two eyes which is not the case for deepfakes images. Nirkin et al. [9] prove this by a deep recognition of hair texture, ears and neck position. Wang et al. [10] use the whole face region to show artifacts in the generated images. Nowadays, physiological and physical-based detection faces a big problem; this is due to the sophisticated evolution of content generation methods, which allow to create images with less visible traces of imperfection to the eye, which makes physiological and physical detection approaches more convoluted. In this work, we are interested in the detection of deepfakes by taking as reference the eyes. We focus our work on the eyes for the simple reason that the eye is one of the organs of the human face which gathers in it elements having a regular and perfectly circular geometrical shape such as the iris and the pupil. Figure 2 shows the parts of the eyes of a human face and presents the details of the parts of the eyes having regular geometrical shapes.

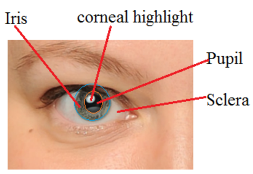


Fig. 2. Image of the eye and its different parts

Our method is based on two existing methods.The real motivation is that the first method is very limited and the second is a complement to the first, applied to normalized color images, thus eliminating a large percentage of false positives and maintaining the rapid increase in the detection rate.

The deepfakes detector generated by GAN is composed of 4 main steps: (1) before any process, the image goes through a face detection model that locates and extracts any human face present in the image. (2) Then these faces go through the pupil shape detector, which is based on a physiological hypothesis showing that pupils have a shape close to the circle or ellipse in the eye of a real face depending on the position of the face in a plane and the angle from which the photo was taken. This is not the case for images generated by StyleGAN2 [11], being the most sophisticated and accurate method of face generation. Images generated by StyleGAN2 most often have a common artifact which is the irregular shape of the pupils. Figure 3 is described as follows: the first two iris on the left represent those of the eyes of a real human face and the two images on the right represent iris of the eyes of a synthetic human face. It is noted that the pupils of the synthesized eyes marked by the interrupted red have an unconventional form and not recognized in geometries. But the main limitation of the previous hypothesis is that

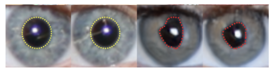


Fig. 3. Pupil samples of the eyes of a real person (left) and of a face synthesized by GAN (right). [3]

when a real person has a visual disorder, the pupil is dilated and takes another shape classified as irregular. In this case, most of these real images will be classified as false when they are not. This is the reason why we have extended this model to have a better detection rate. Then (3) the whole image classified as false is then collected and transformed into gray level to allow the detector to have more accuracy with images being under a precise standard and almost identical whatever the race, the brightness on the face to name but these details. Finally, (4) the images being normalized. This brings us to the third step of our detector, which is checking the corneal reflections of both eyes. the principle is as follows: when a real

Fig. 4. Corneal reflections for a real image (left) and an image generated by the GAN (right). [4]

person looks at an object lit by a light source, the two corneas of both eyes are supposed to reflect the same things in the eyes by respecting certain conditions such as: the light source must be neither too close nor too far from the eyes, the image must be taken in portrait to perfectly present the two eyes, if two lines are drawn perpendicular to the diameter of the iris, these lines must be perfectly parallel. The content generated by GAN does not respect this hypothesis, perhaps because it retains the same characteristics as the actual images selected for learning GAN for image generation. This difference is evident in Figure 4. This figure represents on the left the corneas of the two eyes of a real human face which shows a great resemblance to the naked eye of the observed light sources and the right image takes back the corneas of the two eyes of a synthesized image which unlike the real eyes reflections of images that are unrelated. For some reasons, some false images are classified as true because they have the characteristics described above, so it is here that the last step of our pipeline (4) comes in which is to calculate the ratio of the pixels of the left iris to the right iris. A very striking observation is that the pupils in the eyes of real images are perfectly circular, and the ratio of the pixels in the shape of these pupils to the nearest circle to that pupil is close to 0, whereas for the pupils in the eyes of GAN-synthesized faces, this same ratio is slightly elevated, with perfect inconsistency in classifying false images as true. Our method of detecting the images synthesized by GAN allows us not only to recognize that a false image is false, but also to classify the real images as true with better precision in both cases.

II. Related works

A. GAN: generation of human faces

During the last decade, several GAN architectures have been proposed such as PGGAN [1], BigGAN [12], StyleGAN [2], StyleGAN2 [11], etc. for the synthesis and creation of images of people so realist. Mirsky et al. classify deepfakes into four categories: reconstruction, replacement, editing and synthesis [13]. The first category is a deepfake reconstruction, used to transfer facial expressions from a source individual to a target; for example: change a smile to sadness, or frown, or close the eyes. The second category is the replacement deepfake or face swap used to replace the face of an existing person on an image or video with that of another person without changing anything on the body of the source person. The third category is deepfake editing, which is used to change the age of the source to make it look younger, change the hair color, the color of the clothes or even the race of the person without swapping any element. And the last category is the deepfake of fully generated faces which are people that do not exist in the first place but are generated from artificial intelligence through machine learning in order to generate images that have the same properties as the source images but are not identical to any source image.

B. GAN: detection of fake human faces

The detection of images generated by the GAN can be classified into four categories [14]: detection using deep learning, sensing based on physical data, physiologically based detection and detection by taking into account human observation.

- Detection using deep learning: this locates the area of interest and, using a deep neural network classifier, classifies false images on one side and true images on the other.
- Detection based on physical data: In 2017, Tero Karras et al. [1] demonstrated that in the eyes of GAN-generated faces, corneal specular reflections are either absent, or appear as a simple white spot without representing the object in the scene. In 2021, Shu Hu et al. [4] showed that these reflections are not identical in both eyes.
- Physiology-based detection: This detection method studies artifacts recognized by human vision to classify GAN-generated and real images such as symmetry of facial elements, iris color, pupil shapes. One of the most recent physiology-aware deepfake detection methods is that of Hui Guo et al. [3], which classifies real images from synthesized images based on the simple assumption that the pupils of real eyes are supposed to have a smooth circular or elliptical shape. But in the case of synthesized images, this assumption is unproven because GAN has no knowledge of human eye anatomy, specifically pupil geometry and shape, pays no attention to this clue and the resulting images have dilated, irregularly shaped pupils. Another recent method is that of Xue et al. [15], which exposes GAN-generated images using GLFNet (global and local facial features) [15]. GLFNet is a two-level detection method: a partial detection level based on one or more features, and a global detection level based on detection features such as iris color, pupil shape, false tracks in an image.
- Detection by taking human observation into account: It's not exactly easy for a human being to detect images generated by GAN, since the more the years go by, the more sophisticated GAN models become. In 2021, Federica Lago et al. [16] conducted an experiment to measure a human being's ability to recognize a false image by scrolling through several false and several true images. The results show that it is very difficult to recognize them over time and as the GANs evolve.

C. Normalisation of grayscale images

Fig. 5. Examples of color images (left) transformed to grayscale (right).

Image processing is a growing field because images are widely used in this era and must first be processed to have experiments resulting in excellent results. The processing applied to images are diverse and we can cite: image compression, image segmentation, gray level transformation. Color images are very diverse, depending on race, ethnicity, brightness applied to them, age, sex, etc., so it is difficult to apply detection algorithms to such images and have good results. The ideal was to normalize these images in grayscale in order to obtain all similar images with the same characteristics. An image is converted to grayscale for several reasons: to reduce the light density of an image, or for aesthetic reasons. The conversion to grayscale makes it possible to reduce the three dimensions of colors applied to an image into a single one size [17]. It is all the more true that transforming an image into grayscale also returns a loss of information, but these images present less detail compared to color images and are more likely to produce better results.

III. Method

Our work to detect deepfakes using the face is inspired by the fact that the images generated by GAN present very real and much more perceptible artifacts such as the shape of the pupil and the reflection of the cornea of the eye for a person looking at a lighted image. These two artifacts are the basis of this work because for a real existing person, the pupil of the eye is supposed to be round or elliptical and the two corneas of the eye should reflect the same thing, which is not the case for the images generated by the GAN. The proposed method is briefly shown in 6.

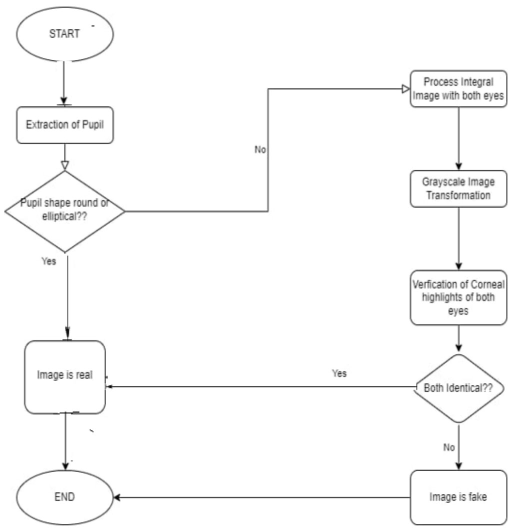


Fig. 6. Flowchart of the method used.

We start by a face detection tool to identify any human face in the input image because this method is only applicable to images of people. Then, thanks to the set of reference points, the eyes are extracted and the areas of interest are located: here it is the pupil on one hand and the cornea on the other.

The boundaries of the curves are then analyzed to check whether the pupil has a regular or irregular shape, by parametrically fitting the shape to an ellipse following RMS (Root Mean Squared) optimization. So, if the pupil has an irregular shape, the second element, the cornea, is targeted. This phase consists of checking the light reflections from the cornea of both eyes. To do this, a complete image of the face of the person looking at a previously tested light source is recovered. The light traces reflected by the two corneas are compared to check whether or not they are identical. In the case of images generated by GAN, this inconsistency is not checked, and the corneal reflections are therefore not identical. A more detailed description of the false face detection process is given in the following paragraphs.

A. *Verification of the shape of the pupil*

We use the method [3] to locate and approximate the shape of the pupil. In this paragraph we will explain the method: The Dlib method [20] is used to detect the face and then extract the 68 landmarks to localize our area of interest which is the eyes in order to segment the pupil. The model used here is EyeCool [21], which is the one used to extract the segmented pupil masks with all contours. It is equipped with a boundary attention block in order to adjust its accuracy to highlight only the boundaries of the object of interest which is the outside of the pupil in order to better perceive the irregularity of its shape. The pupil mask is then fitted to an ellipse by the least squares method. The Boundary IoU (BIoU) method [22], which is a distance metric, is then used to give a percentage match between the irregularity of the pupil shape and the fitted ellipse. The BIoU is most often used in image segmentation with very high boundary sensitivities. BIoU evaluates the IoU of the pixels between the contours of the real mask and the mask created by the fitted ellipse. The BIoU evaluation interval is [0;1]. In other words, when the BIoU value is above the average, it means that the pupil is very close to the ellipse fit and therefore that we are dealing with an existing image, but a smaller value suggests that we should pay attention to the image because it may be false or real, depending on the characteristics of the eyes. The major limitation of this approach is that it is possible to have a high false positive rate because for eyes with certain diseases, it is possible that the pupils of the eyes are irregular as shown in the Figure 7. The figure 7 presents examples of real eyes classified as fake. The first two images on the left present the eyes of two existing people caught in the real face data sets. The two right images represent the irises as well as the pupils of the diseased eyes presented in [18]. For all the images that come out with irregular pupils, we need to be sure if these images are real or fake which gives way to the second step of our pipeline which is to test the corneal specular reflections to see if they are identical.

B. *Verification of the cornea of both eyes*

The work of [4] uses as a clue the elements reflected by the horn of the eye to prove that an image is real or not. To avoid errors in the prediction of results, we have

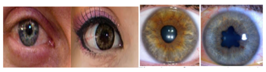


Fig. 7. Examples of images classified as false but true on the left and Examples of diseased and infected eyes from [18] the right

opted for grayscale images for the simple reason that when an image is too bright, it is difficult to accurately identify the reflection of the cornea. This is because a color image has too much detail represented and there is a problem of image uniformity because depending on the race, there may or may not be certain details visible characteristics.

The image having undergone some readjustment can pass to the verification. In the image of a real person taken from a camera,, the reflections of the cornea of the two eyes are identical because they are the result of a same source of light if and only if we take into account certain parameters that we will enumerate hereafter: it is necessary that the two irises are in the same direction and are well in evidence. Under the conditions of a good portrait image, the reflections of the horns of both eyes are almost identical. These conditions are among others: (1) both eyes are well in evidence and if we draw two lines passing both through the center of the pupil, these two lines are perfectly parallel. (2) The two eyes must not be close to the light source so that no matter where the lighted objects in the scene are, they are reflected in the eyes. After extracting the desired area of interest, the limbus of the cornea which represents the circular shape in the eye is extracted and then a Canny edge detector and the Hough transform are applied to better visualize the cornea. Finally, the corneal reflections are extracted from the adaptive image thresholding method [19]. Thus, the IoU score in the range [0; 1] represents a high value if there is similarity and a low value if there is no similarity.

IV. EXPERIMENTS

The proposed method is tested on two datasets: one dataset of true images and another of false images. The dataset of images of people's faces that exist is FFHQ [2]: FFHQ is a dataset with 70,000 images of very good quality at a resolution of 1024 × 1024. It includes a large number of image varieties according to age, ethnic group, race, and image backgrounds. It also includes images of people with accessories such as braces, glasses of any kind, hats, contact lenses etc. For the false image datasets, they are obtained from the StyleGAN2 method [11]: it is a method of image generation based on the GAN and more and that puts more emphasis on the images compared to the GAN with less imperfections visible to the eye. The method has been tested on 1500 real images and 1000 fake images.

We will only present the images classified as negative in the first part of our pipeline, i.e., the images where the pupil of the eye does not have a regular shape and where the IoU score is low that we pass in the second part of our pipeline to check if the image is real. Figure 8 shows the experiments on the images of real people in color as well as the distribution of the different IoU scores in the two cases. It is easy to notice that when the IoU score is low the result can be interpreted by the fact that the image is not real. It is also noticed that the color image contains too much detail of the limuria, so our pipeline has difficulty in predicting the right results correctly. That's why we have opted for grayscale images.

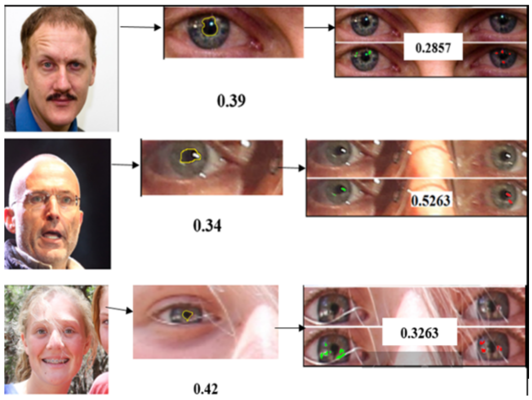
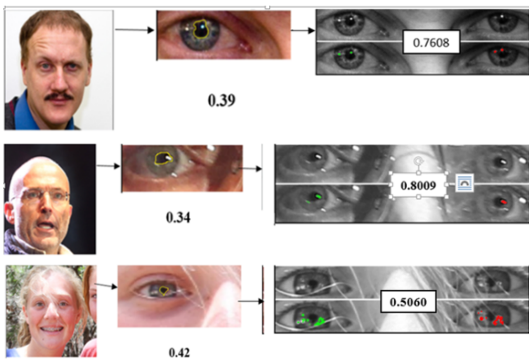


Fig. 8. Experience on images of real persons: color images

The Figure 9 shows examples of results for real images and images generated by the GAN in grayscale as well as the IoU score distribution for both parts of our pipeline. We notice that the grayscale image processing attenuates the light on the face so it is easier to get better results. We notice that not all images with dilated pupils and non-regular shapes are indeed images generated by the GAN. The results of our experiment show that the association of iris shape and corneal image helps a lot to detect the images generated by the GAN with a percentage of success of 0.97 on 1 and a false positive rate of 0.3. A false positive rate is the percentage of deepfake images that the method classifies as real images.

Real images



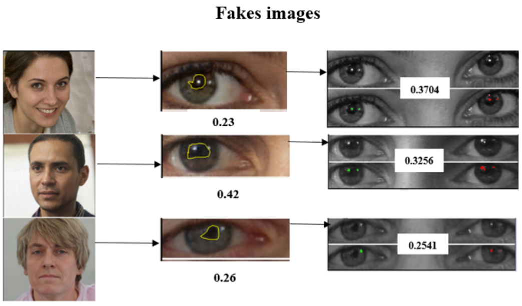


Fig. 9. Experience on images of real persons: grayscale images

A. *Comparison with existing methods*

In short, we tried to compare our method with certain state-of-the-art methods by the AUC (Area under the ROC (receiver operating characteristic) curve), which proves the effectiveness of our method. We selected four methods [3], [4], [7], [15] that give different AUC scores. Hu et al. [4] and Guo et al. [3] for their experiences, took the actual images from the FFHQ datasets and generated images synthesized by the StyleGAN2 method obtained by http://thispersondoesnotexist.com. Ziyu et al [15] used two datasets, FFHQ and CelebA, for real images and two methods of synthesizing false images, StyleGAN2 and ProGAN. The AUC of the method of Hu et al. [4] is 0.94, that of Guo et al. [3] is 0.91 and finally the experience of Ziyu et al [15] has the following results: for 1000 images from the FFHQ dataset of the real images and 1000 images obtained from the StyleGAN2 method have an AUC of 0.96, and for 1000 images obtained from the data set of the real images CelebA and 1000 false images obtained by the ProGAN method they obtain a UAC of 0.88.

TABLE I
COMPARISON OF THE MOST RECENT DEEPFAKES EXPOSURE METHODS FROM THE AUC.

Method	*Real images*	*Fake images*	*AUC*
Hu et al. [4]	***500(FFHQ)***	***500(StyleGAN2)***	***0.94***
Guo et al. [3]	***1000(FFHQ)***	***1000(StyleGAN2)***	***0.91***
Yang et al. [7]	***50 000(CelebA)***	***25 000(ProGAN))***	***0.94***
Xue et al. [15]	***1000(FFHQ)***	***1000(StyleGAN2)***	***0.96***
Xue et al. [15]	***1000(CelebA)***	***1000(ProGAN))***	***0.88***
Our method	***1000(FFHQ)***	***1000(StyleGAN2)***	***0.97***

V. CONCLUSIONS

In this paper, we present a method to expose the images generated by the GAN. It is a two-level detection method, which is to test first the regularity of the pupil shape, because real people's eyes are supposed to have a regular pupil shape, either a circle or an ellipse, depending on the angle at which the image is captured. However, this assertion is often wrong because of the anomalies and diseases that can affect the eyes. Therefore for all images whose pupil does not have a regular shape it is necessary to submit them to the second level of our model which is to check if the horns of the two eyes are almost identical. For this we have first normalized the images in gray levels to have identical images on the shape regardless of the race and also for the light on the image attenuates to make the detection more precise. After this treatment on the images, we have compared the reflections of the two corneas from the IoU score that allows us to better interpret the results. For future work, we will add other elements of the face to improve our detection technique.

REFERENCES

[1] Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)

[2] Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401–4410 (2019)

[3] Guo, H., Hu, S., Wang, X., Chang, M.-C., Lyu, S.: Eyes tell all: Irregular pupil shapes reveal gan-generated faces. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2904–2908 (2022). IEEE

[4] Hu, S., Li, Y., Lyu, S.: Exposing gan-generated faces using inconsistent corneal specular highlights. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2500–2504 (2021). IEEE

[5] Marra, F., Gragnaniello, D., Verdoliva, L., Poggi, G.: Do gans leave artificial fingerprints? In: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 506–511 (2019). IEEE

[6] Wang, S.-Y., Wang, O., Zhang, R., Owens, A., Efros, A.A.: Cnn-generated images are surprisingly easy to spot... for now. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8695–8704 (2020)

[7] Yang, X., Li, Y., Qi, H., Lyu, S.: Exposing gan-synthesized faces using landmark locations. In: Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, pp. 113–118 (2019)

[8] Matern, F., Riess, C., Stamminger, M.: Exploiting visual artifacts to expose deepfakes and face manipulations. In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 83–92 (2019). IEEE

[9] Nirkin, Y., Wolf, L., Keller, Y., Hassner, T.: Deepfake detection based on discrepancies between faces and their context. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(10), 6111–6121 (2021)

[10] Wang, J., Wu, Z., Ouyang, W., Han, X., Chen, J., Jiang, Y.-G., Li, S.-N.: M2tr: Multi-modal multi-scale transformers for deepfake detection. In: Proceedings of the 2022 International Conference on Multimedia Retrieval, pp. 615–623 (2022)

[11] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyziang and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8110-8119 (2020)

[12] Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018)

[13] Mirsky, Y., Lee, W.: The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR) 54(1), 1–41 (2021)

[14] Wang, X., Guo, H., Hu, S., Chang, M.-C., Lyu, S.: Gan-generated faces detection: A survey and new perspectives. arXiv preprint arXiv:2202.07145 (2022)

[15] Xue, Z., Jiang, X., Liu, Q., Wei, Z.: Global–local facial fusion based gan generated fake face detection. Sensors 23(2), 616 (2023)

[16] Lago, F., Pasquini, C., B¨ohme, R., Dumont, H., Goffaux, V., Boato, G.: More real than real: A study on human visual perception of synthetic faces [applications corner]. IEEE Signal Processing Magazine 39(1), 109–116 (2021)

[17] Saravanan, C.: Color image to grayscale image conversion. In: 2010 Second International Conference on Computer Engineering and Applications, vol. 2, pp.196–199 (2010). IEEE
[18] Ramlee, R., Ramli, A., Noh, Z.: Pupil segmentation of abnormal eye using image enhancement in spatial domain. In: IOP Conference Series: Materials Science and Engineering, vol. 210, p. 012031 (2017). IOP Publishing
[19] Yen, J.-C., Chang, F.-J., Chang, S.: A new criterion for automatic multilevel thresholding. IEEE Transactions on Image Processing 4(3), 370–378 (1995)
[20] King, D.E.: Dlib-ml: A machine learning toolkit. The Journal of Machine Learning Research 10, 1755–1758 (2009)
[21] Wang, C., Wang, Y., Zhang, K., Muhammad, J., Lu, T., Zhang, Q., Tian, Q., He, Z., Sun, Z., Zhang, Y., et al.: Nir iris challenge evaluation in non-cooperative environments: Segmentation and localization. In: 2021 IEEE International Joint Conference on Biometrics (IJCB), pp. 1–10 (2021). IEEE
[22] Bowen Cheng, Ross Girshick, Piotr Doll´ ar, Alexander C. Berg, and Alexander Kirillov, "Boundary IoU: Improving object-centric image segmentation evaluation," in CVPR, 2021.