Detecting GAN-generated face morphs using human iris characteristics

ALEXANDRU HESSON, University of Twente, The Netherlands
a.hesson@student.utwente.nl

Abstract

Face morphing is a computer graphic technique that merges the features of two individuals into a single composite image and it finds applications in diverse fields such as entertainment and cosmetic surgery simulation. However, it also poses a potential threat in the form of identity document forgery. This research focuses on the detection of morphed facial images generated by Generative Adversarial Networks (GANs) by analysing the characteristics of the iris. In genuine human irises, the shape is predominantly circular, whereas morphed images may exhibit extreme deformations or deviate from the expected roundness. To address this issue, a program capable of scanning facial images and extracting the iris contour for radius samples was developed to be used in evaluating the iris roundness. A dataset comprising 44 authentic irises and 44 morphed irises was compiled for experimenting, ensuring diversity by including individuals from various ethnic backgrounds. The findings from this study showed that using iris detection alone is not accurate enough to reliably detect morphs, but it can be added to a multi-facet approach to improve the accuracy of already existing morphing detection systems.

Keywords: Face morphing, GAN-generated, morphing detection, iris scan.

1. Introduction

1.1. Face morphing

Face morphing has emerged as a significant security concern for Face Recognition Systems (FRS), particularly in the context of Automatic Border Control (ABC) gates that rely on Machine Readable Documents (eMRTD). These systems play a crucial role in ensuring efficient and secure border control procedures by verifying the identity of individuals using their biometric information. However, the increasing sophistication of face morphing techniques poses a serious threat to the integrity of such systems.

One specific type of attack, known as the "magic passport" attack[1], has raised considerable alarm within the field. This attack takes advantage of a vulnerability present in certain countries where individuals are allowed to select their own ID or passport photo by presenting authorities with a printed picture of their choice. Exploiting this loophole, individuals with criminal records seeking international travel may attempt a magic passport attack.

The process of a magic passport attack begins with an accomplice who bears a resemblance to the target individual. The accomplice generates a morphed image by combining their facial features with those of the target. The resulting morph exhibits traits from both individuals, creating a convincing fraudulent identity. Subsequently, the accomplice reports their passport as lost, and during the issuance of a new passport, they present the printed morphed image as their legitimate photo.

The success of a face morph relies on its ability to deceive human experts and appear ICAO compliant, adhering to the standards set by the International Civil Aviation Organization. Poorly created morphs are immediately noticeable due to visible artifacts or improperly blended textures. However, these issues can be addressed by skilled professionals who apply manual post-processing techniques to refine the morphed image. On the other hand, well-created morphs are extremely challenging to detect, even for experts in the field.

The task of differentiating between a morphed image and a genuine one on a passport becomes increasingly difficult considering that eMRTD photos are typically small, measuring only 3.5x4.5cm[18]. The limited size and resolution of these images make it harder to discern the intricacies and inconsistencies that may indicate a morphed image. A study[2] conducted in this domain has revealed that matching unfamiliar and unknown faces is highly prone to errors, further increasing the vulnerability of incorporating morphed images into passports.

In response to these growing security concerns, numerous studies have been conducted to detect morphed images using various techniques. These techniques include analysing micro-texture variations[3], texture noise[4] and comparing artifacts between eyes[5], among others. However, existing state-of-the-art methods for detecting facial morphs have limitations, as their training models are often based on datasets that do not adequately represent real-world scenarios[5].

To address this issue, this study aims to explore the detection of morphed images generated by Generative Adversarial Networks (GANs) by examining the characteristics of the iris. The human iris is predominantly circular, except in rare cases resulting from conditions like coloboma, which occur in less than one in every 10,000 births. By leveraging the roundness feature of the iris, this study seeks to develop more robust and reliable techniques for detecting morphed images in the context of eMRTD authentication.

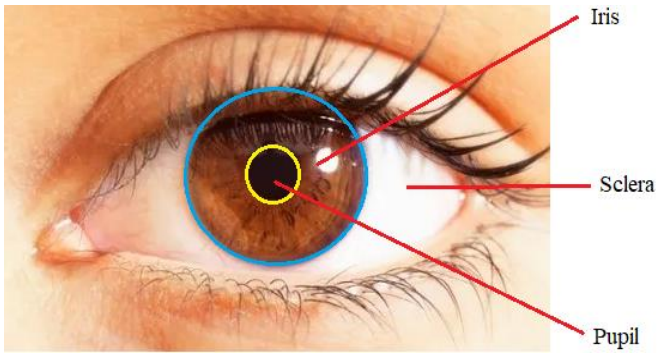


Figure 1: The human iris

Enhancing the detection capabilities against morphed images is of great importance to maintain the integrity and security of face recognition systems, particularly in the context of ABC gates and eMRTD verification. By addressing the limitations of existing methods and exploring innovative approaches, this research aims to contribute to the development of more effective countermeasures and bolster the overall security of border control systems worldwide.

1.2. GAN vs landmark-based morphs.

Traditional landmark-based morphing algorithms rely on the precise alignment of facial landmarks from two images and the subsequent blending of corresponding regions to create a morphed image. In contrast, GAN-generated morphs employ a deep learning-based technique using Generative Adversarial Networks (GANs)[13] that consists of a generator network and a discriminator network. The generator network synthesizes new images based on the learned patterns and distributions of a training dataset, while the discriminator network distinguishes between real and generated images. This adversarial process encourages the generator to produce more realistic and convincing images over time.

Research[13] has indicated that, in a digital environment, GAN-generated morphs do not pose a significantly greater threat to Face Recognition Systems (FRS) compared to traditional morphing algorithms. However, it is crucial to recognize that GAN-generated morphs introduce potential risks in the print-scan process. This process involves printing a morphed image and subsequently scanning it to create a physical document such as a passport or ID card.

Figure 2: Examples of real and morphed irises

1.3. Research Question

This study will focus on answering the following research question:

To what level of accuracy can the roundness characteristic of the human iris be used by image recognition systems to detect GAN-generated image morphs?

2. Dataset Creation

To gather data for this study, two distinct datasets were compiled and tested: one comprising real irises and another consisting of morphed irises. The iris images were analysed to investigate the distribution of iris radiuses and compare the average values between the real and morphed irises. Further details on the methodology employed in this study will be discussed in the Methodology chapter.

The dataset utilized in this study for morphed irises was sourced from the FRLL-Morphs dataset, which encompasses a collection of morphed faces obtained from the publicly available dataset provided by Face Research London Lab[6]. A meticulous manual cropping process was undertaken to extract the irises from the corresponding images, ensuring a diverse representation of ethnicities and genders within the dataset.

In the case of acquiring real iris samples, a comprehensive approach was adopted. Images encompassing a wide spectrum of ethnicities and genders were procured from

reputable free stock photo platforms, specifically Pexels[7] and Shutterstock[8]. For consistency the resolution of the selected images was as close as possible to the morphed images resolution. The irises within these images were consistently and manually cropped following the same rigorous methodology applied to the morphed iris dataset.

The resultant dataset employed in this study consisted of 44 morphed irises and 44 real irises, providing a robust foundation for thorough data analysis and comparative examinations.

3. Related Work

An investigation has been undertaken to detect GAN-generated faces by analysing irregularities present in the shapes of pupils [15]. However, it is important to note that this study specifically targeted the identification of generated faces rather than facial morphs. Nevertheless, the findings of this study have yielded valuable insights into the creation process of GAN-generated images, as well as their distinctive characteristics.

Pioneering work in the field of GAN-generated morphs includes MorGan[3], which generates morphed images of 120x120 pixels. However, these images do not comply with the standards set by the International Civil Aviation Organization (ICAO), rendering them impractical for real-world attacks[13]. In recent advancements, methods like StyleGAN[14] have emerged, allowing for the generation of high-resolution images at 1024x1024 pixels while maintaining ICAO compliance. This opens possibilities for GAN-generated morphs to be utilized in sophisticated attacks such as magic-passport scenarios.

IrisParseNet [17] presents a compelling solution for iris segmentation, employing a deep-learning algorithm to accurately capture the contours of the iris. This innovative approach enables comprehensive iris segmentation even in challenging conditions, such as non-cooperative environments or when working with low-resolution and noisy images but trying to replicate or implement this sophisticated approach in this study proved to be highly unfeasible and fell well outside of the scope of this research. An additional study [16], which focused on iris contour detection, achieved success in accurately isolating the iris contour as well as the pupil through the utilization of geodesic active contours. However, the methodology employed in that particular study was also not feasible for implementation within the context of this present study.

The book [18] served as a valuable reference for accessing in-depth information regarding the morphing process, its intricacies, and the diverse number of morphing detection systems. This publication offers a comprehensive starting point for acquiring detailed insights into morphing, including a detailed understanding of its complexities and the wide range of methodologies employed for detecting morphed images.

4. Methodology

This section will cover the process of extracting the iris, drawing its contour, and getting the radius samples.

The present study was conducted utilizing the programming language Python[9], along with the libraries Numpy[10] and Matplotlib[11]. Numpy was selected due to its robust capabilities in handling extensive multi-dimensional arrays and matrices, as well as its diverse collection of arithmetic operations designed to operate on such arrays. Matplotlib, on the other hand, was employed for the purpose of visualizing data through the creation of graphs and histograms, thereby enhancing the analysis of this study.

Furthermore, for the image processing component OpenCV[12], or Open Source Computer Vision Library, was specifically chosen. This decision was based on several factors, including the user-friendly nature of OpenCV, its exceptional integration with the Python programming language, and its provision of built-in functions, such as automatic contour detection and the Hough Circle Transformation, which greatly facilitate image analysis and manipulation.

The initial step in extracting the iris contour entails converting the original image to grayscale and applying a blurring technique. Grayscaling is employed to optimize the performance of the built-in contour detection functionality in OpenCV, while blurring is implemented to diminish noise and enhance the accuracy of contour detection.

Subsequently, the iris region is detected by utilizing the Hough Circle Transformation function implemented by OpenCV, a built-in feature that intelligently positions a perfect circle over the iris. However, recognizing the necessity for more comprehensive information about the iris contour, an additional step is taken. By expanding the radius of the resultant circle, a larger circular area is encompassed, encompassing the iris as well as its surrounding region, thus being possible to detect the iris contour more accurately. Although exact contour extraction of the visible iris posed challenges due to my limitations in image processing and OpenCV expertise, a pragmatic alternative was implemented. Consequently, the bottom half of the iris, which typically encompasses a significant portion of the iris contour, was selectively

extracted, yielding satisfactory outcomes for most iris images.

After the iris region detection process is finished the resulting semi-circular or elliptical iris shape is merged using a bitwise OR operation with a white image of equal resolution to the original image. The application of thresholding to this resulting image, in conjunction with the built-in contour detection method provided by OpenCV, facilitates the visualization of distinct contours. These contours are then placed onto the original image, effectively illustrating the extracted iris contour for visual analysis.

To obtain radius samples, the pixel distance is measured from the center of the eye to the contour edges utilizing the formula:

$$((x1 - y1)^2 - (x2 - y2)^2)^{1/2}$$

x1 is the x-axis coordinate of the center, x2 is the x-axis of the contour point and y1,y2 being the y-axis coordinates of the center and contour point respectively. The center coordinates are determined by taking the Hough Circle center coordinates and adjusting the X-axis value by dividing the radius of the eye by two. However, due to inherent limitations, no suitable correction mechanism could be found for the Y-axis, as the iris is almost never entirely visible within the captured image.
The quantity of radius samples obtained from the real iris pictures exhibits variation due to variations in resolution, typically ranging between 70 and 120 samples, but the number of samples taken from morphed irises was more inconsistent as even though the images had a constant resolution of 1024x1024, some images were blurrier than others. The radius samples necessitate filtering to account for reflections present in the iris, as demonstrated in Figure 3.

To address this issue, a condition was introduced as a filtering mechanism. The condition examines each radius and removes it from the list if it is smaller than 80% of the average radius. The threshold of 80% was selected after careful consideration, as it demonstrated effectiveness in preventing interference with genuine contour detection between the iris and the sclera, while concurrently mitigating the detection of reflections. It should be noted that reflections occurring along the border between the iris and the sclera were challenging to filter out, albeit their occurrence was infrequent. The iris region was divided into 6 regions for a more thorough analysis of the iris contour. Initially there was a hypothesis suggesting the potential existence of regional iris deformations, but later in the study this proved to not be true or useful.

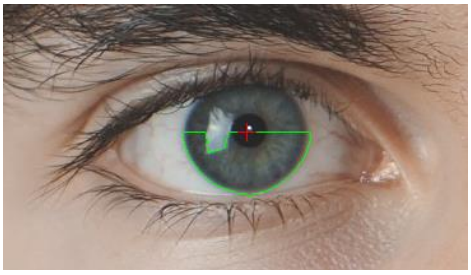


Figure 3: showing a reflexion getting picked up by the contour detection.

Figure 4: showing the steps required to extract the iris contour.

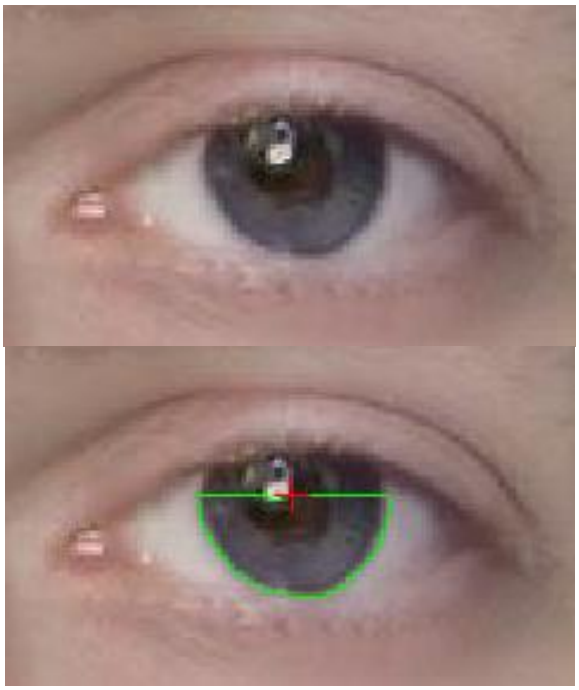
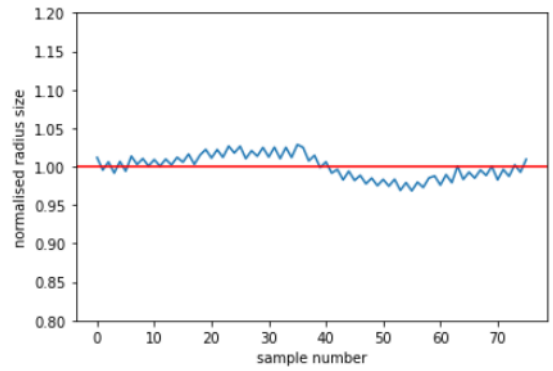
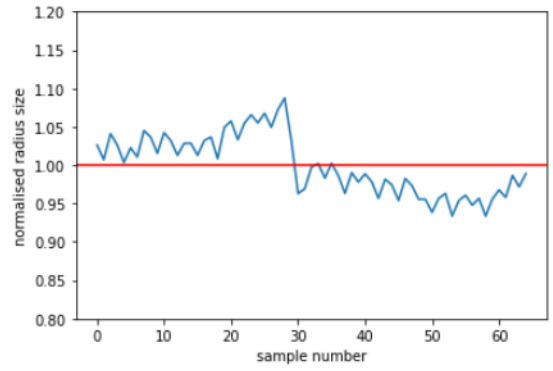


Figure 5: Deformed morphed iris

The resulting radius samples will be normalized for consistency by dividing them to the average, then plotted on a graph for visual inspection. The average will be recalculated and plotted as a red line on the graph.
To calculate how spread the samples are compared to the average the relative standard deviation was chosen as a unit of measurement. The relative standard deviation was chosen instead of the normal standard deviation because it gives the deviation value in percentages, instead of pixels. This was chosen because there is no global pixel reference between images since the resolution varies. More images of real and morphed irises along with their radius distribution graphs will be showed in the appendix.



relative standard deviation: 1.58%



relative standard deviation: 3.88%

Figure 5: plotted graphs of radius measurements. First graph is drawn from a real iris and the second graph from a deformed morphed iris.

On a perfect circle the relative standard deviation of the radius samples will always be 0%, due to their identical values, but the human iris is not perfectly round and there is some degree of error to my contour detection algorithm. This error comes from the following points:

-The X-axis coordinate of the circle's center can occasionally be a floating-point number (e.g., 57.5), requiring rounding. While applying pixel segmentation to the image can mitigate this issue, it presents a significant challenge within the limited scope and timeframe of this project.
-When the iris colour is blue or light green, the contrast between the iris and the sclera may not be substantial enough to facilitate effective separation through thresholding. Consequently, the contour detection process may encroach slightly into the sclera region.
-The resolution of the images poses the most substantial error factor. After cropping the eye region from the facial picture, the resulting resolution may be insufficient to accurately depict the contour with complete precision.

Some error is to be expected in the radius measuring, so even real irises have a degree of relative standard deviation.

5. Experiments and Results

The first experiment conducted was comparing the average relative standard deviation between the real iris dataset and the morphed iris dataset. The result was that the average relative std for the real iris dataset was 1.71% and for the morphed irises 2.43%. This shows that on average a morphed iris will have 42.1% more relative std compared to a real iris, so the next experiment was to plot the distribution of the relative std on a histogram to see the overlap between the averages of the real and morphed irises.

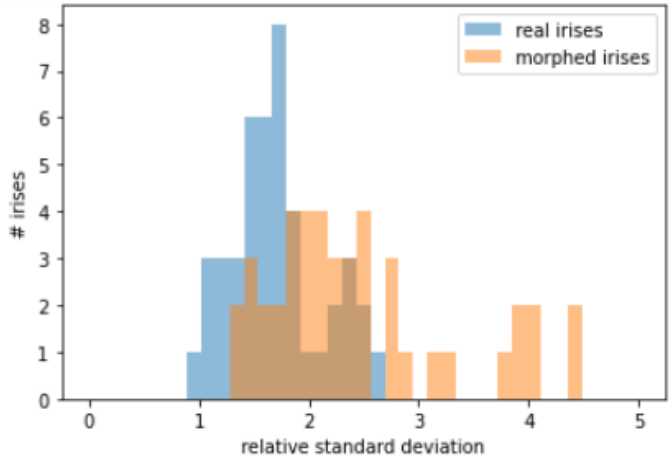


Figure 6: Histogram showing the relative standard distribution for all irises.

The findings derived from the histogram analysis revealed that among the 44 examined morphed irises, 13 out of 44 (29.5%) exhibited a relative standard deviation higher than that observed in all real irises. Additionally, it was observed that 7 out of the 44 real irises (15.9%) displayed a relative standard deviation lower than the minimum relative standard deviation observed in the dataset of morphed irises. This left 68/88 (77.3%) of total irises having an overlapping relative standard deviation.
This outcome was anticipated, given that a significant proportion of the morphed images exhibit irises with relatively circular shapes, while only a few display noticeably distorted irises. Most authentic irises are concentrated around the mean of 1.71%, displaying a slight positive skewed to the right distribution. In contrast, the distribution of morphed irises appears more dispersed, with the majority clustered around the 2% mark. However, notable outliers exist within this group, representing morphed irises that are distinctly deformed. After those results, some manual labelling was done on the morphed irises, labelling them in two categories: visibly deformed and not visibly deformed, and another histogram was plotted.

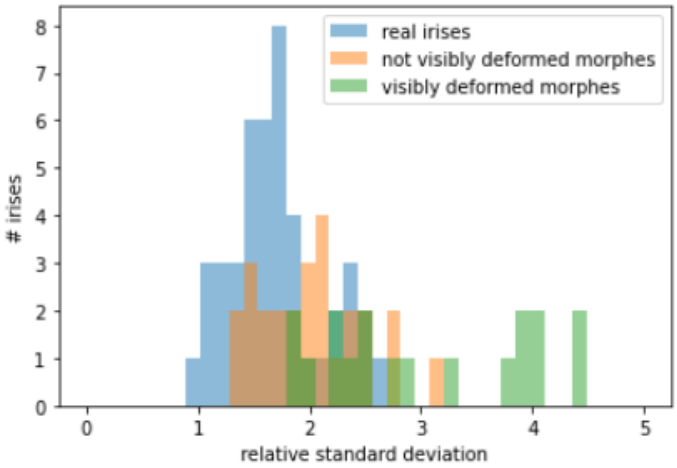


Figure 7: Histogram showing the relative std of visibly and non-visibly distorted morphs compared to real irises.

This labelling resulted in the average relative standard deviation to drop for the not visibly deformed morphed irises to 2.01%, and the average of the visibly deformed morphed irises to increase to 3.03%. Furthermore, 10/18(55.55%) visibly deformed morphs had a relative std higher than all real irises. The accuracy can be improved by refining the iris segmentation algorithm further, more detail about this will be discussed in the Future Work chapter.

6. Future Work

Future work in this field includes several improvements and expansions to enhance the contour detection algorithm and broaden the capabilities of iris detection and analysis. These advancements aim to address specific scenarios, increase accuracy, and strengthen the overall validity of the study.

One area for improvement is refining the contour detection algorithm to perform effectively in scenarios where the iris has lighter colours, such as blue or green. The algorithm could be optimized to handle variations in colour and texture, enabling accurate segmentation and contour extraction even in challenging cases.
Another avenue for enhancement is introducing pixel segmentation to account for instances when the X-axis coordinates of the iris center have floating-point values. By incorporating precise pixel-level segmentation, the algorithm's accuracy can be increased, allowing for more precise measurement of the iris radiuses.

Expanding the scope of the iris detection program is also a notable future improvement. Currently, the program is limited to covering only half of the visible iris structure. To overcome this limitation, an enhancement could involve applying a mask over the sclera region and intersecting it with the iris region. This approach would effectively remove the eyebrows and eyelids from the analysis, allowing for a more comprehensive and accurate detection of the complete visible iris structure.

To further fortify the study's validity, increasing the sample size for the datasets is an important consideration. A larger dataset would provide a broader range of examples and variations, enabling the validation of the algorithms. A larger sample size would also help in evaluating the algorithm's robustness.

By addressing these areas of improvement, future work aims to enhance the accuracy, versatility, and reliability of the iris detection and analysis system. These advancements will contribute to the field of biometric identification and authentication, increasing the overall security and effectiveness of iris-based recognition systems in various applications and domains.

7. Conclusions

To answer the research question, detecting facial morphs solely through iris detection is challenging due to the presence of morphed irises that still exhibit a relatively round shape, the accuracy can be significantly enhanced by combining iris detection with other morphing detection methods.

Based on the data obtained from the utilized datasets, the iris detection method alone would demonstrate an accuracy of approximately 29.5% in correctly identifying morphed images while avoiding misclassification of real images. This accuracy is attributed to the fact that 13 out of the total 44 morphed images displayed a higher iris radius relative standard deviation compared to all real irises. If the iris is visibly deformed the accuracy increases to 55.55%.

While this accuracy rate might appear relatively low, especially compared to other studies[3,4,5,18] where the accuracy is usually over 90%, it is important to note that detecting facial morphs solely through iris detection is indeed challenging due to the complexities of morphed irises that can still maintain a round shape. However, these findings suggest that incorporating iris detection as part of a multi-faceted approach can yield promising results in enhancing the overall accuracy of morph detection.

To conclude this study, findings highlight the importance of adopting a multi-modal approach to detect facial morphs accurately. By integrating multiple detection methods, researchers can leverage the strengths of each technique, compensating for their respective limitations. This approach has the potential to significantly enhance

the accuracy and reliability of morph detection systems, enabling more robust identification of morphed images and contributing to the security and integrity of face recognition applications.

Acknowledgements

I would like to thank my supervisor, Luuk Spreeuwers, for all the support and guidance he gave me throughout this project, as well as the feedback, and numerous ideas that helped me immensely into finishing this study. I would also like to thank the track chair, Estefanía Talavera Martínez, for ensuring a nice organization of the research project.

References

[1] Ferrara, Matteo & Franco, Annalisa & Maltoni, Davide. (2014). The magic passport. IJCB 2014 - 2014 IEEE/IAPR International Joint Conference on Biometrics. 10.1109/BTAS.2014.6996240.

[2] Robertson DJ et al. (2018) Detecting morphed passport photos: a training and individual differences approach. Cogn Res Princ Implic 3(27)

[3] Ramachandra, Raghavendra & Raja, Kiran & Busch, Christoph. (2016). Detecting Morphed Face Images. 10.1109/BTAS.2016.7791169.

[4] Fu, Tao & Xia, Ming & Yang, Gaobo. (2022). Detecting GAN-generated face images via hybrid texture and sensor noise based features. Multimedia Tools and Applications. 81. 10.1007/s11042-022-12661-1.

[5] H. Guo, S. Hu, X. Wang, M. -C. Chang and S. Lyu, (2022). "Robust Attentive Deep Neural Network for Detecting GAN-Generated Faces," in IEEE Access, vol. 10, pp. 32574-32583

[6]https://www.idiap.ch/en/dataset/frll-morphs (June 2023)

[7] https://www.pexels.com

[8]https://www.shutterstock.com/search/human-face

[9] Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

[10] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020). DOI: 10.1038/s41586-020-2649-2.

[11] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", (2007). Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95.

[12] Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.

[13] Venkatesh, Sushma & Zhang, Haoyu & Ramachandra, Raghavendra & Raja, Kiran & Damer, Naser & Busch, Christoph. (2020). Can GAN Generated Morphs Threaten Face Recognition Systems Equally as Landmark Based Morphs? -- Vulnerability and Detection.

[14] Karras, T., Laine, S., & Aila, T. (2018). A Style-Based Generator Architecture for Generative Adversarial Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4396-4405.

[15] H. Guo, S. Hu, X. Wang, M. -C. Chang and S. Lyu, "Eyes Tell All: Irregular Pupil Shapes Reveal GAN-Generated Faces," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 2904-2908, doi: 10.1109/ICASSP43922.2022.9746597.

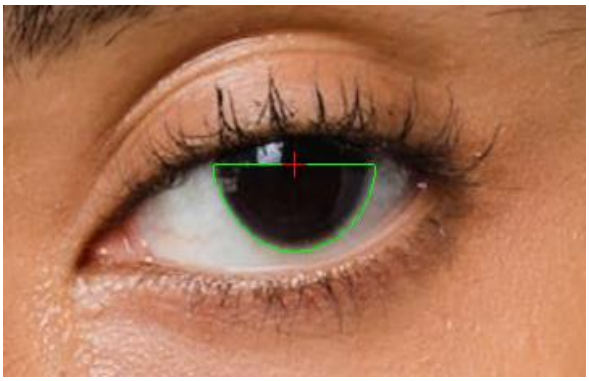
[16] Shah, Saransh & Ross, Arun. (2010). Iris Segmentation Using Geodesic Active Contours. Information Forensics and Security, IEEE Transactions on. 4. 824 - 836. 10.1109/TIFS.2009.2033225.

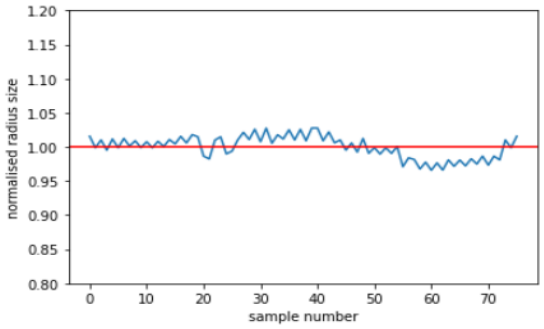
[17] Wang, Caiyong & Muhammad, Jawad & Wang, Yunlong & He, Zhaofeng & Sun, Zhenan. (2020). Towards Complete and Accurate Iris Segmentation Using Deep Multi-Task Attention Network for Non-Cooperative Iris Recognition. IEEE Transactions on Information Forensics and Security. PP. 1-1. 10.1109/TIFS.2020.2980791.

[18] Rathgeb C, Tolosana R, Vera-Rodriguez R, Busch C (2022) "Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks" (pp. 331-349)

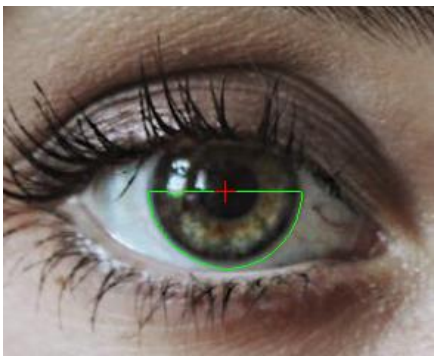
Appendixes

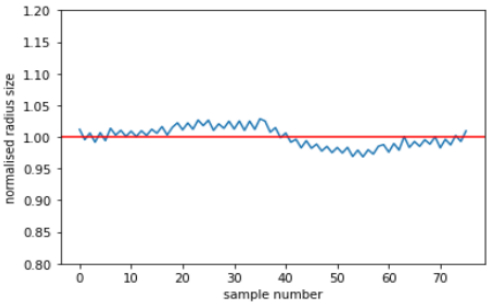
Appendix A – Real irises



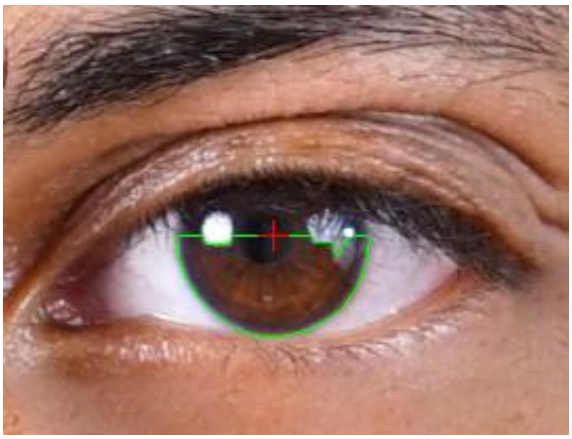


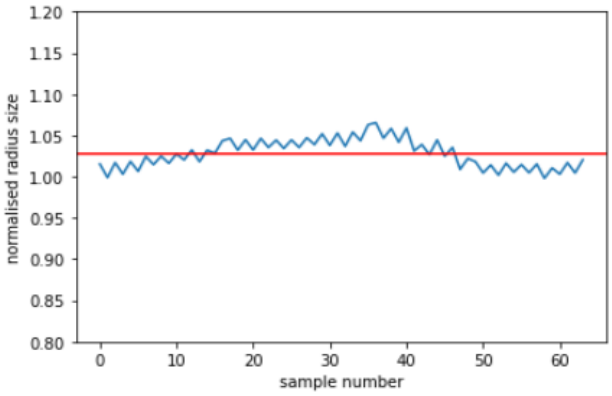
relative standard deviation: 1.645%



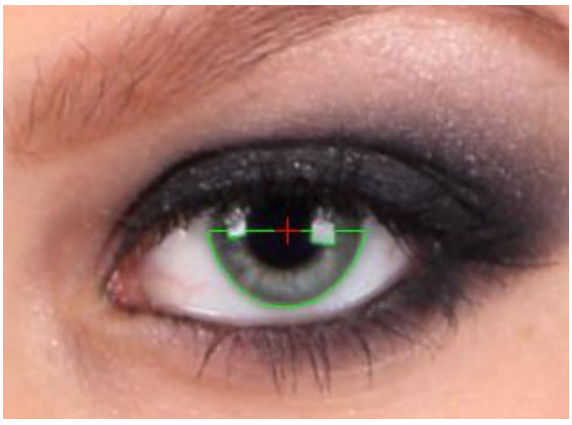


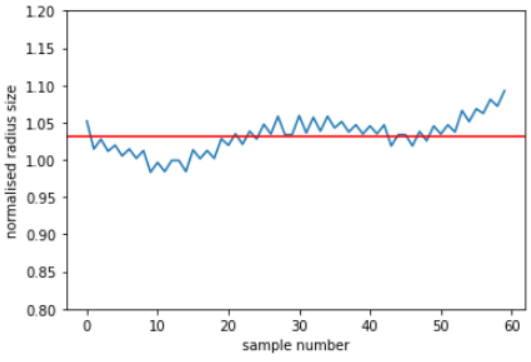
relative standard deviation: 1.58%



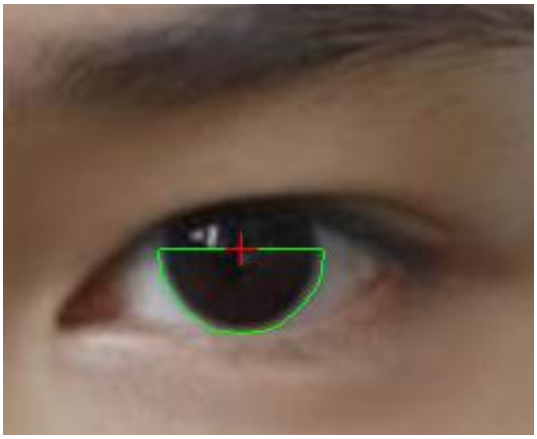


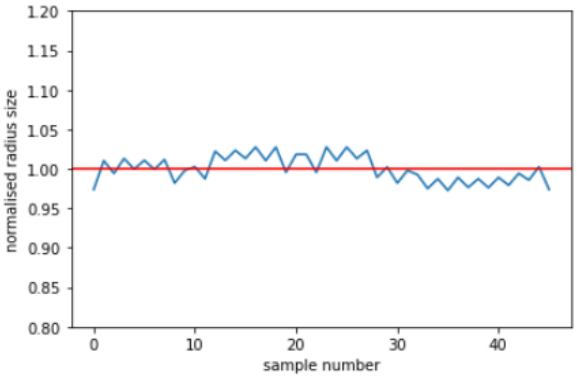
relative standard deviation: 1.678%



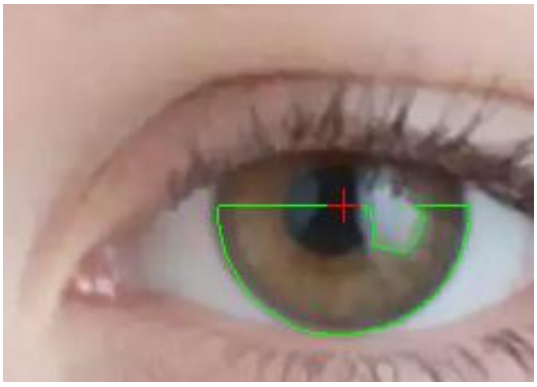


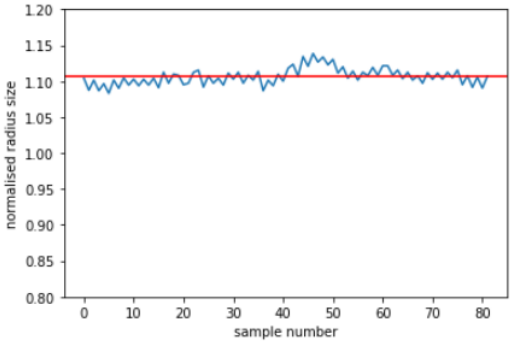
relative standard deviation: 2.286%





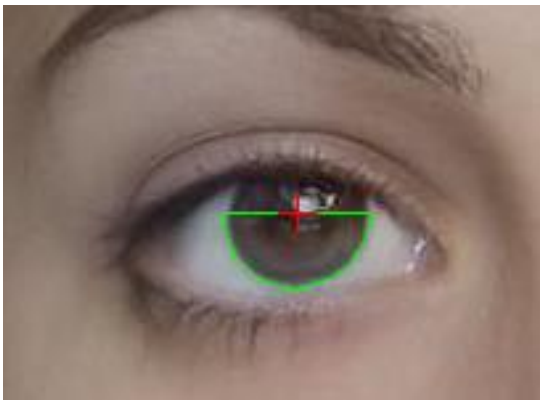
relative standard deviation: 1.662%

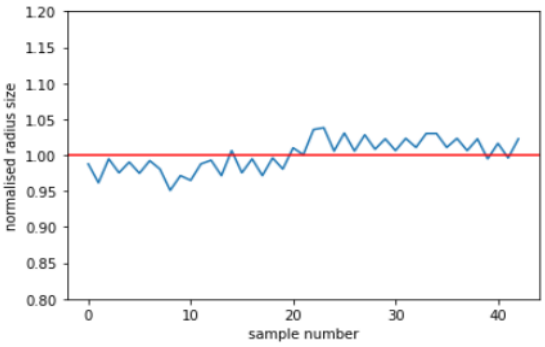




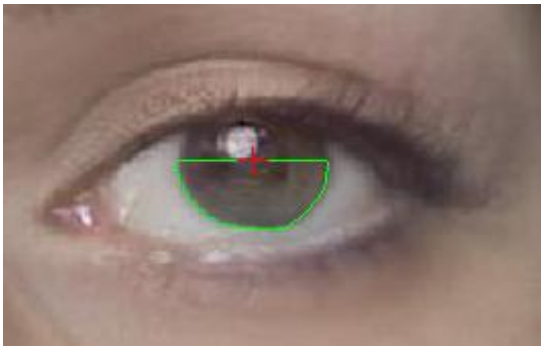
relative standard deviation: 1.028%

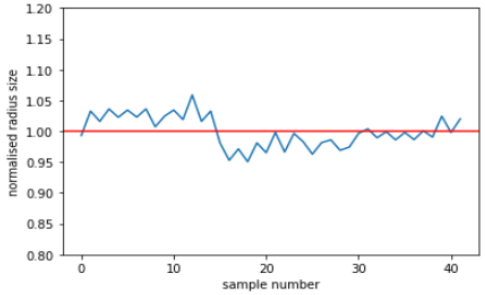
Appendix B – Morphed irises



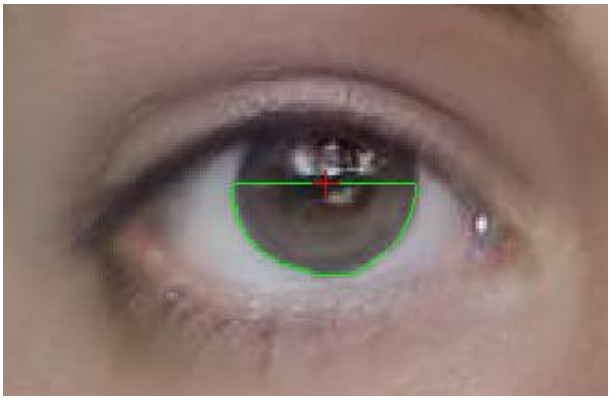


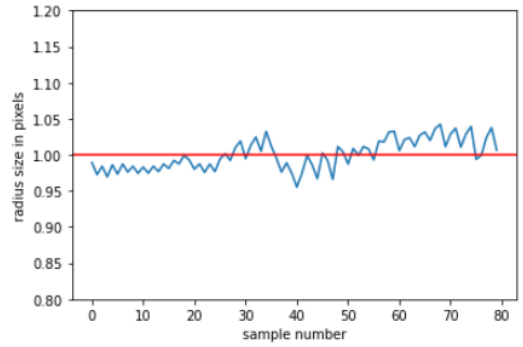
relative standard deviation: 2.184%



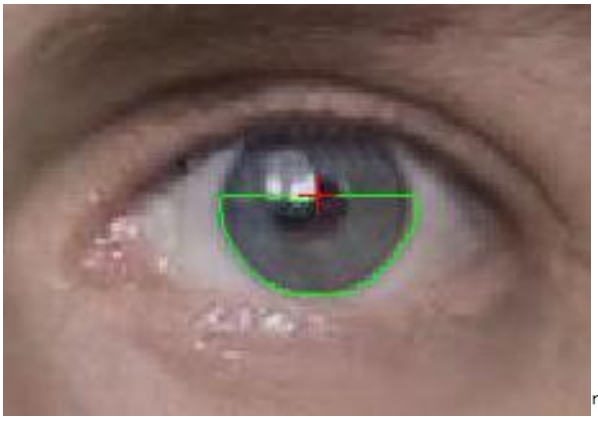


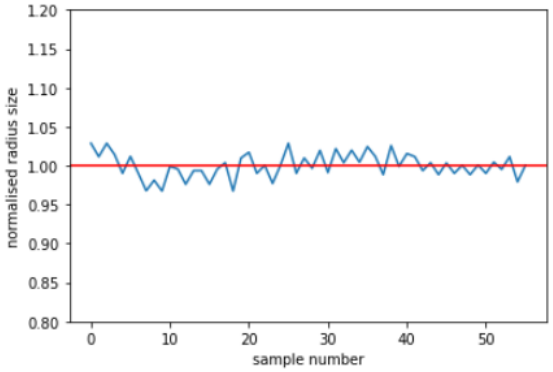
relative standard deviation: 2.544%



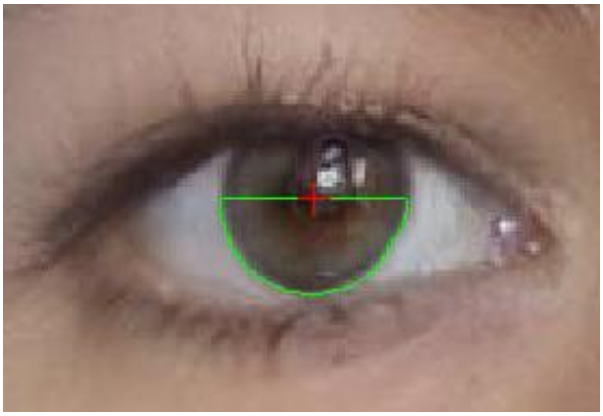


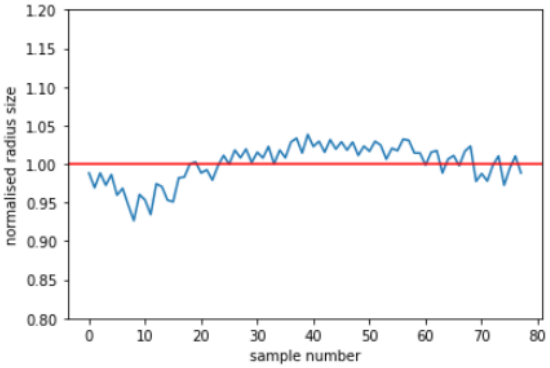
relative standard deviation: 2.084%



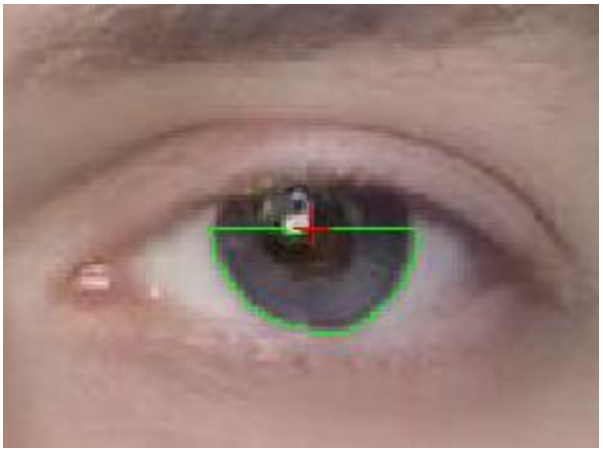


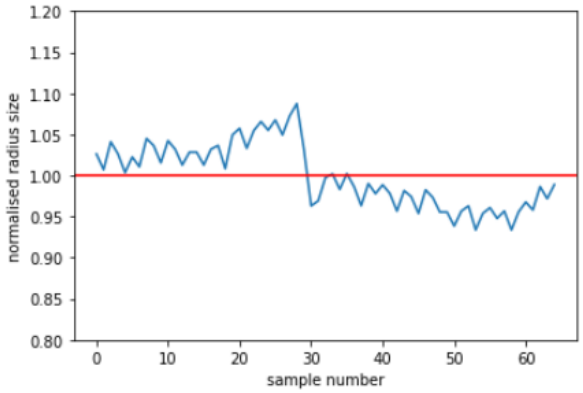
relative standard deviation: 1.555%





relative standard deviation: 2.493%





relative standard deviation: 3.88%