Synthetic Iris Presentation Attack using iDCGAN

Naman Kohli[1], Daksha Yadav[1], Mayank Vatsa[1,2], Richa Singh[1,2], Afzel Noore[2]
[1]West Virginia University, USA, [2]IIIT Delhi, India
{naman.kohli, daksha.yadav, afzel.noore}@mail.wvu.edu, {mayank, rsingh}@iiitd.ac.in

Abstract

Reliability and accuracy nature of iris biometric modality has prompted its large-scale deployment for critical applications such as border control and national ID projects. The extensive growth of iris recognition systems has raised apprehensions about susceptibility of these systems to various attacks. In the past, researchers have examined the impact of various iris presentation attacks such as textured contact lenses and print attacks. In this research, we present a novel presentation attack using deep learning based synthetic iris generation. Utilizing the generative capability of deep convolutional generative adversarial networks and iris quality metrics, we propose a new framework, named as iDCGAN (iris deep convolutional generative adversarial network) for generating realistic appearing synthetic iris images. We demonstrate the effect of these synthetically generated iris images as presentation attack on iris recognition by using a commercial system. The state-of-the-art presentation attack detection framework, DESIST is utilized to analyze if it can discriminate these synthetically generated iris images from real images. The experimental results illustrate that mitigating the proposed synthetic presentation attack is of paramount importance.

1. Introduction

Beauty lies in the iris of the beholder! Some of the iris images in Figure 1 are not real iris images. Can you identify which ones have been generated synthetically?

Ratha et. al. [18] presented several avenues of attack on a biometric system and suggested different steps to mitigate such attacks. One of these avenues is through presentation attacks at sensor level which can be used both for identity impersonation and identity evasion. The other potential point of attack in a biometric system is the transmission channel between the sensing device and the feature extraction module [18]. A man-in-the-middle attack on this channel can be utilized to replace the original image with a new synthetic image before the template extraction process.

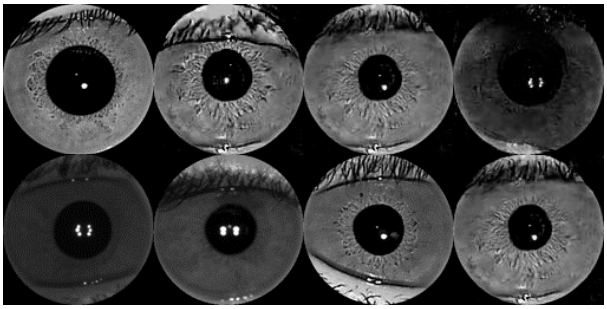


Figure 1: A mixture of real and synthetic iris images generated from the proposed iDCGAN framework are shown above. We encourage the readers to identify which of these iris images are real and synthetic. The solution is shown in Figure 7 on page 6.

The consequences of such an attack maybe wide-ranging as an individual may enrol with different identities and avail facilities associated with the unique ID multiple times.

Presentation attacks on iris modality such as textured contact lenses [12, 24], synthetically generated iris [6], and print attacks [8] have been explored in the literature. The idea of generating synthetic iris images was initially introduced by Cui et al. [3] with the intention of increasing the number of available iris images for developing iris recognition algorithms. They employed principal component analysis and super-resolution techniques to create new images for iris synthesis. Shah and Ross [19] employed Markov Random Field to generate initial texture of the iris images followed by embedding iris features such as radial and concentric furrows to create the final synthetic iris image. Zuo et al. [26] developed an anatomy based model to create new irises similar to real-world iris images. Galbally et al. [6] reconstructed synthetic iris images from the feature template to successfully match the original genuine iris image. Figure 2 shows sample synthetic iris images from Synthetic DataBase (SDB) by Galbally et al. [6]. It is seen that these images do not resemble real iris images and appear *fake*.

In this paper, we propose a new iris presentation attack by synthesizing iris images through deep convolutional

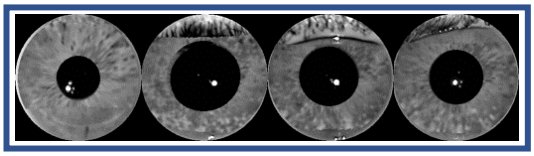


Figure 2: Sample images from Synthetic DataBase [6].

generative adversarial network. Recently, improvements in techniques such as generative adversarial networks [7] and variational autoencoders [11] have provided a breakthrough in generative modeling. These approaches have paved the path for generating realistic looking synthetic images for different applications. In this research, we have proposed a novel synthetic iris image generation method using generative adversarial network and demonstrated that it can attack iris recognition systems. The major contributions of this paper are:

1. A novel domain specific generative adversarial network (GAN) named as iDCGAN for generating synthetic iris images is proposed. We adapt deep convolutional generative adversarial network by utilizing iris quality assessment for synthesizing *realistic looking* iris images.
2. Analysis is performed using quality score distributions of real and synthetically generated iris images to understand the effectiveness of the proposed approach. We also demonstrate that synthetically generated iris images can be used to attack existing iris recognition systems.
3. Evaluation using state-of-the-art iris presentation attack detection algorithm is performed to ascertain its efficacy in distinguishing these synthetically generated images from real images.

2. Generative Adversarial Network for Iris Image Generation

In this research, we adapt generative adversarial network for synthesizing realistic iris images to propose iris Deep Convolutional Generative Adversarial Network (iDCGAN). Figure 3 shows the steps involved in the proposed approach.

2.1. Generative Adversarial Network

Goodfellow et al. [7] introduced the concept of generative adversarial networks (GANs) where the generative model is pitted against an adversarial *discriminator* to generate representations which cannot be differentiated by the discriminator. The aim of the *generator* is to learn the probability distribution of the input data perfectly enough to *fool* the discriminator.

Let $\mathbf{x}$ be the input data which has a true probability distribution $p(\mathbf{x})$. Let $\mathcal{G}$ be the generative network which takes an input latent vector $\mathbf{z}$, drawn from a noisy probability distribution $p_{noise}(\mathbf{z})$ and outputs a new image $\bar{\mathbf{x}}$. Then, the discriminator network $\mathcal{D}$ has to discern if the input image, randomly chosen from $\mathbf{x}$ or $\bar{\mathbf{x}}$, is generated from the true probability distribution $p(\mathbf{x})$ or not. The two models are trained using a minimax objective and the loss function L is shown in Eq. 1.

$$L = \min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}[log(\mathcal{D}(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}\sim p_{noise}(\mathbf{z})}[1 - log(\mathcal{G}(\mathcal{D}(\mathbf{z}))] \quad (1)$$

A number of variants of GANs have been introduced such as conditional GANs [16], Laplacian GANs [5], and InfoGANs [2]. These variants have been successfully utilized in image inpainting [25], style transfer [23], and super-resolution [15] applications. Recently, Shrivastava et al. proposed SimGAN [20] which uses a refiner network to improve appearance of synthetically generated eye images to make them indistinguishable from real eye images.

2.2. Proposed iDCGAN for Iris Image Synthesis

Radford et al. [17] introduced deep convolutional generative adversarial networks (DCGAN) for unsupervised learning of features by utilizing convolutional neural networks as the generator and discriminator network. They also applied constraints on architectural topology of convolutional neural networks in the generator and discriminator networks for stable training. Specifically, pooling functions were replaced with strided convolutions which allowed the resultant network to learn its own spatial upsampling. Additionally, the fully connected layers at the top of convolutional neural networks were removed and batch normalization was utilized for improving model stability by normalizing each unit to have zero mean and unit variance.

In this paper, we propose an extension to DCGAN by utilizing domain (iris) specific knowledge. The new generative adversarial network is termed as iDCGAN (iris Deep Convolutional Generative Adversarial Network). Similar to the idea of conditional GANs [16], it uses auxiliary information of iris quality to improve the performance of both discriminator and generator deep convolutional networks.

In any iris recognition system, iris image quality assessment is an integral step as the quality of iris images can greatly impact the performance of iris recognition. It has been ascertained that different artifacts such as occlusion, off-gaze direction, motion blurriness, and specular reflection can affect iris recognition performance [10, 21]. Thus, incorporating quality metrics in generative adversarial network can improve the synthesis process. Eq. 2 shows the objective function of the proposed iDCGAN framework.

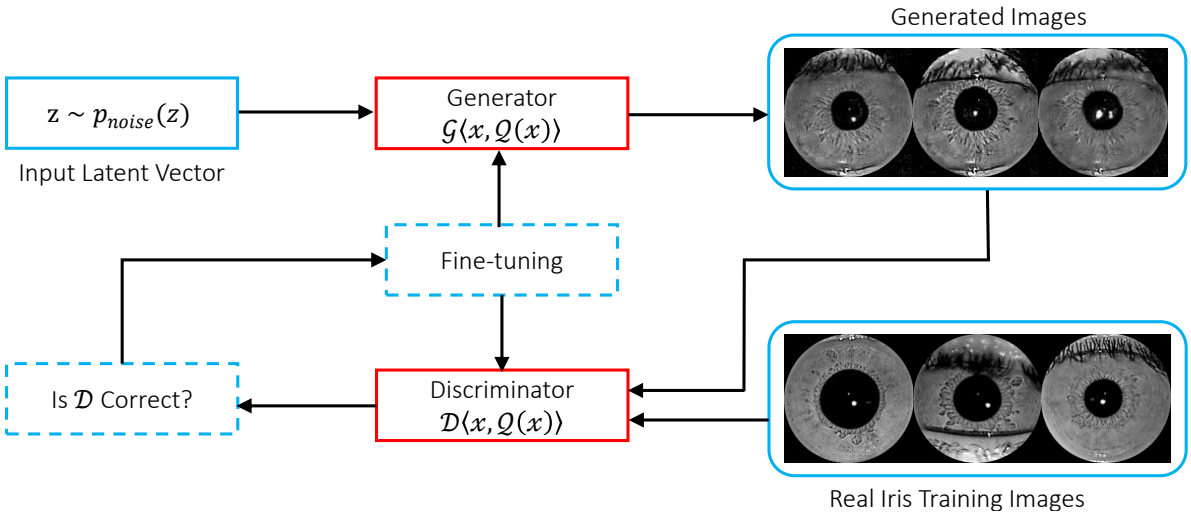


Figure 3: Illustrating the proposed iDCGAN framework for generating synthetic iris images.

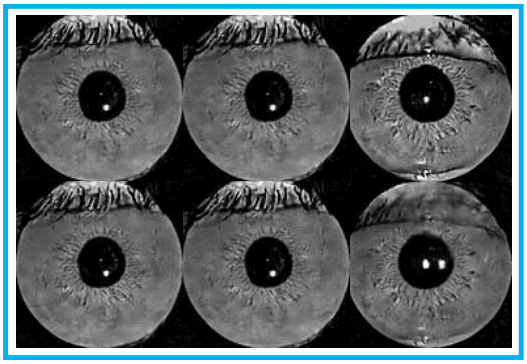


Figure 4: Sample synthetic iris images generated from the proposed iDCGAN framework.

$$L = \min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}[\log(\mathcal{D}(\langle \mathbf{x}, Q(\mathbf{x})\rangle))] + \mathbb{E}_{\mathbf{z}\sim p_{noise}(\mathbf{z})}[1 - \log(\mathcal{G}(\mathcal{D}(\langle \mathbf{z}, Q(\mathbf{z})\rangle)))] \quad (2)$$

where, $Q(\mathbf{x})$ is a quality evaluating function that takes an input iris image and assigns a corresponding quality score. Thus, in the proposed iDCGAN framework the generator network $\mathcal{G}$, spawns new images of iris conditioned on high quality scores.

The input latent vector is generated from a noisy distribution $p(\mathbf{z})$. This is provided as input to the generator network, where the generator generates iris images according to the learned representations. Quality assessment of the iris images created by the generator $\mathcal{G}$ is performed. The quality of the iris images in the first quartile are removed from the set to be passed to the discriminator network $\mathcal{D}$. Similar to the above step, the real iris image input to the discriminator network $\mathcal{D}$ is filtered such that the training set contains iris images whose quality scores are above the first quartile. The new samples are continously generated to train the proposed iDCGAN generator and discriminator. Figure 4 showcases sample iris images generated from the proposed iDCGAN framework.

2.3. Implementation Details

Three existing real iris image databases are utilized and combined together to form the training set for the proposed iDCGAN framework:

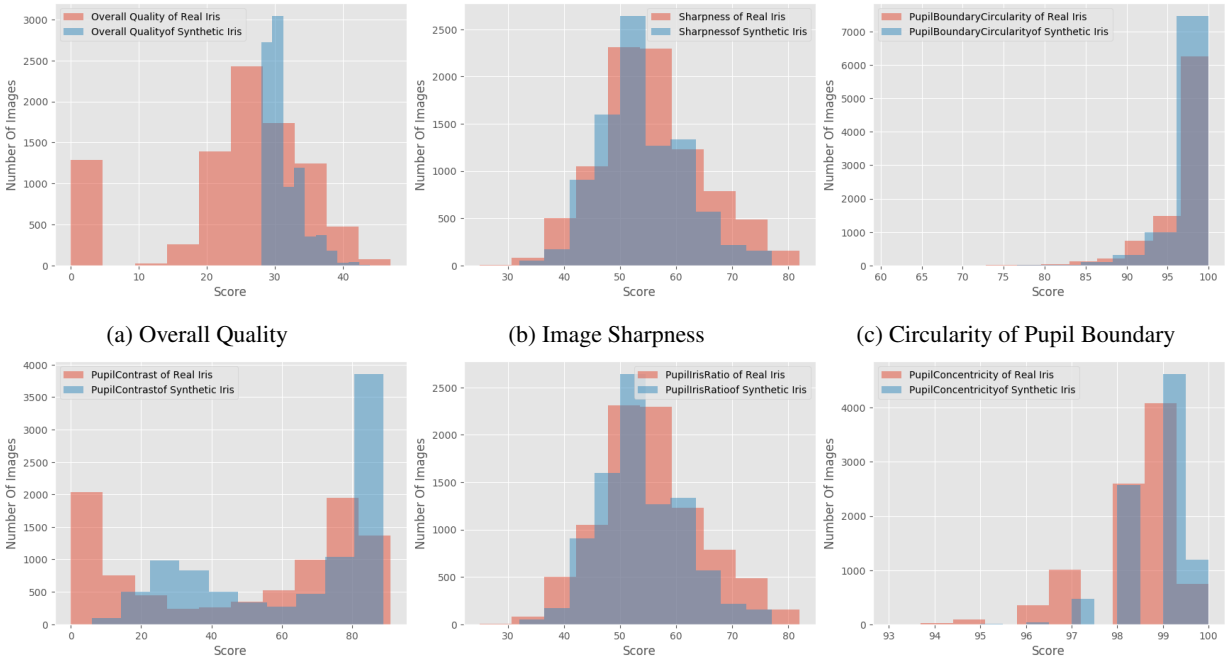
IIITD Contact Lens Database [24] This database consists of iris images of 101 subjects. The database includes iris images of subjects with and without contact lens. For training the proposed iDCGAN, only the real images (without contact lens) belonging to these subjects are chosen.

IIT Delhi Iris Database [14] This database consists of real iris images pertaining to 224 subjects.

MultiSensor Iris Database [13] Iris images of 547 subjects collected in multiple sessions are utilized for training the proposed iDCGAN framework.

The input iris images are segmented so that only the iris and pupil regions are considered as input to the iDCGAN framework. The framework is implemented in Python language utilizing the TensorFlow library[1]. Both the generator and discriminator networks are deep convolutional neural networks. The discriminator network consists of four convolutional layers with kernel size of 5×5 and strides of 2, batch normalization and leaky rectified units.The generator network consists of four strided transposed convolutional layers with kernel size of 5×5 and strides of 2, batch

[1] https://www.tensorflow.org



(a) Overall Quality (b) Image Sharpness (c) Circularity of Pupil Boundary

(d) Pupil Contrast (e) Pupil-to-Iris Ratio (f) Pupil Concentricity

Figure 5: Distribution of various quality metrics highlights the overlap between real and synthetically generated iris images.

normalization and rectified units. The size of the final synthetic iris images is 128×128. A learning rate of 0.0002 and Adam optimizer are utilized to train the proposed iDCGAN.

3. Analysis of Synthetically Generated Iris Images

The synthetic iris images produced by the proposed iDCGAN framework are evaluated with respect to their similarity with real iris images as well as their ability to attack the iris recognition systems. For this, two experiments are conducted which are described below.

3.1. Analysis using Iris Quality Metrics

The iris images generated using the proposed iDCGAN framework are compared with real iris images and are evaluated with respect to different quality score metrics. The quality metrics can evaluate factors such as sharpness of generated images, shape and concentricity of pupil and iris.

3.1.1 Experimental Protocol

The objective of this experiment is to determine the quality of the synthetically generated iris images and compare the quality score distribution with real iris images. Using the combined training set described above, 8,905 real iris images are selected. This is followed by generating an equal number of synthetic iris images using the proposed iDCGAN framework. Bharadwaj et al. [1] described that the quality of iris images can be categorized into image based and biometric modality based quality measures. Using VeriEye, several image specific and biometric modality specific quality scores are computed. These quality score metrics are described in ISO/IEC 29794-6 standards [9]. The following quality score metrics are employed for the analysis purposes:

- Pupil boundary circularity: This parameter represents the circularity of the iris-pupil boundary. It is calculated as $\left(2 * \sqrt{\pi \times \text{pupil area}}\right) / (\text{pupil perimeter})$.
- Pupil contrast: The contrast value at the boundary of iris and pupil is an important parameter for successful iris segmentation. It is computed as the mean of differences in grayscale values at left and right end of iris-pupil boundary.
- Pupil-iris ratio: This quality measure signifies the amount of dilation or constriction in the pupil.
- Pupil concentricity: This parameter measures the corresponding concentricity between the iris and the pupil. It is calculated as $\sqrt{(X_{pupil} - X_{iris})^2 + (Y_{pupil} - Y_{iris})^2}/IrisRadius$ where X and Y represent the coordinates of the iris and pupil.

- Sharpness: The sharpness of the image parameter is examined to understand the magnitude of defocus in the input iris image. This is calculated using Daugman's focus score [4].
- Overall quality: The overall quality score of the iris image represents the comprehensive biometric quality of the presented iris sample. We have utilized output quality score generated from VeriEye.

3.1.2 Results and Analysis

Figure 5 showcases the distributions of the above mentioned quality parameters pertaining to real iris images and synthetically generated iris images. We observe that the quality measurements of the synthetically generated images follow similar trends to the real iris images. The analysis of the quality metrics can be categorized as follows:

Image based Quality: The sharpness score is an image based quality metric. It is observed that there is a significant overlap between the histograms of sharpness observed in real iris images and synthetically generated iris images. The χ^2 distance between the sharpness quality histograms is 1.07 which is relatively low[2]. Similarly, pupil contrast parameter represents contrast difference in a specific region of interest in the image. The χ^2 distance between the pupil contrast histogram is 4.02. It can be observed that the pupil contrast of synthetically generated images is skewed on the higher side as compared to the pupil contrast of real iris images. Thus, larger number of synthetically generated iris images using the proposed iDCGAN framework have higher pupil contrast score as compared to real iris images.

Biometric based Quality: The pupil-iris ratio, pupil boundary circularity, and pupil concentricity are measures of the iris biometric modality. We observe that there is a significant overlap between the distribution of pupil-iris ratio, pupil concentricity and pupil boundary which is also confirmed by the χ^2 distance of 1.07, 0.04 and 0.34, respectively.

Overall Quality: The quality of the synthetically generated iris images is skewed on the higher side and is different from the quality of the real iris images in the combined training set. The generator network in the proposed iDCGAN framework is trained to discard iris images that are not of good quality. Therefore, it has generated high quality synthetic images.

The comparative analysis of these quality score metrics indicates that the synthetically generated iris images very closely resemble the real iris images.

3.2. Synthetic Iris as Presentation Attack

The objective of the proposed iDCGAN framework is to generate iris images which appear *real*. Due to the realistic appearance of these synthetic iris images, they can be used as an attack on any iris recognition system. In this experiment, we utilize VeriEye [22] to examine if a commercial iris recognition matches these synthetic images to real iris images. The results of this experiment are utilized to establish that the output images from the proposed iDCGAN framework can act as an iris presentation attack.

3.2.1 Experimental Setup

The goal of this experiment is to compute iris recognition scores between gallery and probe sets to evaluate the impact of synthetically generated iris as presentation attacks. For this iris recognition experiment, real genuine, real impostor, and synthetic impostor pairs are created using 8,905 real iris images and 8,905 synthetic iris images. The match scores obtained by matching these pairs are analyzed and the results are presented below.

3.2.2 Results and Analysis

These real genuine and synthetic impostor scores are analyzed to observe the impact of synthetically generated iris images on the performance of VeriEye. Upon minimizing the synthetic iris false accept to 0%, we observe that 15.2% of real iris genuine scores are misclassified as impostors. On the other hand, minimizing the real iris false reject to 0% leads to synthetic false accept rate of 67.66%. This showcases that the synthetically generated images adversely affect iris recognition and can pass through the recognition system based on the chosen permissible error threshold.

Interestingly, we observe that all the synthetically generated iris images are encoded by VeriEye and templates are created for every image. A denial of service attack can easily be executed on an iris recognition system by sending such synthetically generated iris images as input. These results validate that the realistic-looking synthetically-generated iris outputs from the proposed iDCGAN framework can be potentially used for iris presentation attack.

4. Iris Presentation Attack Detection on iDCGAN Generated Iris Images

The key results of the previous section illustrate that the synthetically generated iris images from the proposed iDCGAN framework can be effectively deployed in iris presentation attacks. Hence, it is important to develop accurate iris presentation attack detection (PAD) algorithms which can distinguish such synthetic iris images from real iris images. In this section, we present baseline results of state-of-the-art PAD algorithm, DESIST [13].

[2]Lower χ^2 distance values signify very close match.

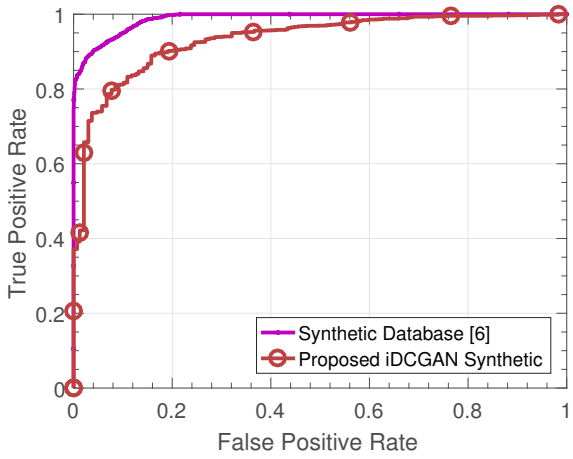


Figure 6: Performance of presentation attack detection using DESIST on images from the Synthetic DataBase [6] and the proposed iDCGAN synthetic images.

4.1. Experimental Protocol

In this experiment, we analyze the performance of DESIST PAD algorithm for detecting synthetically generated iris images. To showcase that the synthetically generated iris images using the proposed iDCGAN framework are strong adversary as compared to existing synthetic iris images, we utilize SDB [6]. SDB comprises 2,100 synthetic iris images. Equal number of real iris images and iris images that are synthetically generated from the iDCGAN approach, are utilized for experimental evaluation. In this experiment, five-fold cross validation is performed with unseen training and testing samples. Multi-order Zernike moments and local binary pattern with variance (LBPV) features are extracted to provide input to the DESIST framework for classifying iris images as real or synthetic using neural network as the classifier.

4.2. Results

The results of the presentation attack detection using DESIST are presented in Figure 6. Iris PAD accuracy on the synthetically generated iris images using the proposed iDCGAN framework is 85.95% with equal error rate (EER) of 14.19%. PAD performance of DESIST on SDB is 92.17% with an EER of 7.09%. We observe that EER by DESIST on SDB is approximately 2 times higher than the EER obtained with iDCGAN generated images. As discussed in the previous sections, the iris image quality scores of the realistic appearing synthetically generated samples are closer to the real-world samples and hence, it is difficult for the DESIST model to discriminate between the samples of the real iris and presentation attack iris classes.

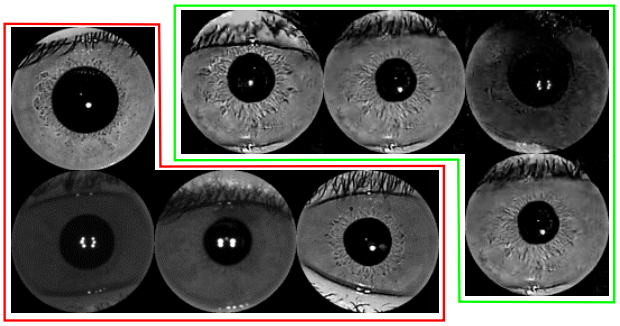


Figure 7: Marked real iris and synthetically generated iris images using the proposed iDCGAN framework. Iris images inside the red border are real iris images and the remaining iris images inside the green border are synthetically generated images.

5. Conclusion

The answer to the question posed in Figure 1 is shown in Figure 7. The iDCGAN framework incorporates iris domain specific knowledge in the form of quality metric to generate high quality iris images. It is observed that the distributions of quality parameters described for a biometric sample for the synthetically generated iris images are similar to that of real iris images, thus, establishing the similarity between real and synthetically generated images. We also demonstrate the probability of a successful presentation attack by utilizing these synthetically generated iris images. Finally, state-of-the-art presentation attack detection framework, DESIST is applied to distinguish synthetically generated iris images from real images. It is observed that the synthetically generated iris images from the iDCGAN framework are more challenging to be detected by DESIST compared to existing synthetic iris database. This paper also highlights the need to develop accurate iris presentation attack detection algorithms that can adapt to newer types of attacks.

6. Acknowledgement

The authors gratefully acknowledge the support of NVIDIA Corporation for the donation of the Tesla K40 GPU for this research.

References

[1] S. Bharadwaj, M. Vatsa, and R. Singh. Biometric quality: a review of fingerprint, iris, and face. *EURASIP Journal on Image and Video Processing*, 2014(1):34, 2014.

[2] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2172–2180, 2016.

[3] J. Cui, Y. Wang, J. Huang, T. Tan, and Z. Sun. An iris image synthesis method based on PCA and super-resolution. In *International Conference on Pattern Recognition*, pages 471–474, 2004.

[4] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[5] E. L. Denton, S. Chintala, R. Fergus, et al. Deep generative image models using a Laplacian pyramid of adversarial networks. In *Advances in Neural Information Processing Systems*, pages 1486–1494, 2015.

[6] J. Galbally, A. Ross, M. Gomez-Barrero, J. Fierrez, and J. Ortega-Garcia. Iris image reconstruction from binary templates: An efficient probabilistic approach based on genetic algorithms. *Computer Vision and Image Understanding*, 117(10):1512 – 1525, 2013.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[8] P. Gupta, S. Behera, M. Vatsa, and R. Singh. On iris spoofing using print attack. In *International Conference on Pattern Recognition*, pages 1681–1686, 2014.

[9] Information Technology - Biometric Sample Quality. Standard ISO/IEC 29794-6:2015 - Part 6 - Iris image data.

[10] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic. Image quality assessment for iris biometric. In *Defense and Security Symposium*, pages 62020D–62020D. International Society for Optics and Photonics, 2006.

[11] D. P. Kingma and M. Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, number 2014, pages 1–14, 2013.

[12] N. Kohli, D. Yadav, M. Vatsa, and R. Singh. Revisiting iris recognition with color cosmetic contact lenses. In *International Conference on Biometrics*, pages 1–7, 2013.

[13] N. Kohli, D. Yadav, M. Vatsa, R. Singh, and A. Noore. Detecting medley of iris spoofing attacks using DESIST. In *IEEE International Conference on Biometrics Theory, Applications and Systems*, pages 1–6, 2016.

[14] A. Kumar and A. Passi. Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recognition*, 43(3):1016 – 1026, 2010.

[15] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. *CoRR*, abs/1609.04802, 2016.

[16] M. Mirza and S. Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.

[17] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015.

[18] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.

[19] S. Shah and A. Ross. Generating synthetic irises by feature agglomeration. In *IEEE International Conference on Image Processing*, pages 317–320, 2006.

[20] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. Learning from simulated and unsupervised images through adversarial training. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–10, 2017.

[21] M. Vatsa, R. Singh, and A. Noore. Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 38(4):1021–1035, 2008.

[22] VeriEye. `http://www.neurotechnology.com/verieye.html`.

[23] X. Wang and A. Gupta. Generative image modeling using style and structure adversarial networks. In *European Conference on Computer Vision*, pages 318–335, 2016.

[24] D. Yadav, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, and K. W. Bowyer. Unraveling the effect of textured contact lenses on iris recognition. *IEEE Transactions on Information Forensics and Security*, 9(5):851–862, 2014.

[25] R. Yeh, C. Chen, T. Lim, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with perceptual and contextual losses. *CoRR*, abs/1607.07539, 2016.

[26] J. Zuo, N. A. Schmid, and X. Chen. On generation and analysis of synthetic iris images. *IEEE Transactions on Information Forensics and Security*, 2(1):77–90, 2007.