Regular paper

Unsupervised Biometric Anti-spoofing using Generative Adversarial Networks

Vishu Gupta[†], Masakatsu Nishigaki[†], and Tetsushi Ohki[†]

[†]Faculty of Informatics, Shizuoka University, Japan
vishu@sec.inf.shizuoka.ac.jp, {nisigaki, ohki}@inf.shizuoka.ac.jp

Abstract - With the advent of new technologies, the methods of presentation attacks as well as the security measures taken against it is diversifying with each passing day and are competing with each other. The imposter can make access to a system illegally by deceiving the security through the use of material containing artificial biometrics traits like a printed photo, display, etc. Therefore, we propose a novel presentation attack detection algorithm which can ensure security against unknown presentation attacks without any prior knowledge of fake samples. Moreover, our proposed algorithm can detect presentation attack with a single static image only. The essential tasks are divided into two parts, creating a smooth manifold of live samples and determining whether the manifolds includes the query image. In this paper, we utilize one class system such as SVM(Support Vector Machine) and DCGAN(Deep Convolutional Generative Adversarial Network) to learn the manifold of live samples. For DCGAN we propose a liveness scoring scheme based on the AnoGAN(Anomaly Generative Adversarial Network) Framework. Based on these, we utilize the proposed method to palm presentation attack detection. Through our experiment, we were able to produce decent results by using palm live/fake image dataset.

Keywords: biometrics, spoofing, presentation attack detection, anomaly detection, generative adversarial networks

1 INTRODUCTION

Along with the development of artificial intelligence and cryptographic technology, a society approaching not only simple tasks but also decision making of people to computers is coming. In such a society, it will become an important requirement to guarantee that the outsourcing was performed by the user's own will, and also to correctly detect it when counterfeiting acts are forged or improperly tampered. It is essential to guarantee the authenticity of the terminal in addition to the authenticity of the terminal itself to satisfy these requirementsṪhe biometric authentication system is drawing attention, which can guarantee the authenticity of the terminal user.

Biometric authentication system (BAS) registers preliminary collected biometric information as a template and verifies whether it belongs to a legitimate user by calculating the similarity with the biometric information acquired at the time of authentication. BAS uses a biometric feature of the person without fear of forgetting, losing, or theft compared to an authentication method using a password or a token. In addition to the advancement of traditional application in fields such as immigration control, ATM, the entry and exit management, recent years, personal use in mobile terminals has been expanding.

On the other hand, biometric information such as faces, sounds, fingerprints, handwriting is difficult to keep secret in daily life. Biometric presentation attack is becoming a significant threat since false biometric information becomes more sophisticated along with the rapid development of sensors, printers, and manufacturing machines.

To develop a BAS that is secure against presentation attacks, demand for designing a robust presentation attack detection(PAD) algorithms which classify an input sample as live or fake is increasing.

Many previous approaches discussed the PAD features which can guarantee security against a specific impersonation attack such as frequency spectrum for printed photo [1], [2], three dimensionalities of live face [3], motion-based feature for video [4] and so on. However, the methods of presentation attacks are diversifying day by day. It is difficult to learn in advance PAD features that can detect all these attacks.

Regarding the problem, PAD algorithms have made it possible to detect various presentation attacks by combining multiclass classifier that solves the classification problem between live and various fake samples such as [5]–[8]. However, these methods still have some issues. At first, it is necessary to obtain not only biometric samples but also a large number of fake samples for each type of presentation attack. Second, the PAD algorithm does not guarantee whether an anomaly sample is classified as a presentation attack. Here we define anomaly sample as a sample that is not included in the samples for training. Note that anomaly sample includes not only samples intended to resemble a live sample but also any synthetic samples since it is sufficiently effective in the registration process. There exists an attack using synthesized input that can impersonate the majority of registered users [9]. Also, attacks that send arbitrary commands to unregistered home interactive speakers by using sounds in the inaudible area [10]. Capturing such attacks with pre-trained PAD features is difficult.

The subject of this paper is to investigate the security against the presentation attack using an anomaly sample as features of biometric information such as the face, palm, etc. differ depending on the modality. Therefore, we utilize DCGAN to perform the estimation of the distribution of biometric information to solve the fundamental problem of making PAD difficult due to the diversification of attacks. Moreover, it is impossible to predict the counterfeit that will be used as an impersonation of the real sample while performing the anti-spoofing using a biometric system. Therefore, by making use of one class system neural network which uses anomaly detection to distinguish fake sample from the true sample, we can make a better system to counter spoofing attacks.

In the experiment conducted in this paper, we used a custom-made database created out of palm images as it was easy to create unknown samples as well as it was found realistic that the attacker may perform the counterfeit attack by using a rubber glove, displayed photo, etc. in an attempt to break into the system.

Additionally, our method relies on a single static image to detect presentation attack. Such a method can also be directly applied to deal with video spoof or be integrated with a video-based palm PAD algorithm for better performance. The main contributions of this work are as follows:

1. We propose a novel Presentation Attack Detection (PAD) algorithm which can be learned only with live samples and guarantee security against an anomaly sample by utilizing GAN based anomaly detection algorithm.

2. Proposed PAD algorithm is evaluated with custom-made database (Custom-Made Database containing live and fake palm samples) and achieved 3% of HTER (Half Total Error Rate) by using a model trained only with live samples.

2 RELATED WORK

All the prior research that has been conducted on Anomaly detection is performed by having to train the system by using both live samples along with fake samples which are used for presentation attack. For all these conducted researches, the core difference lies in the method used to model the real and fake attempts. Prior methodologies based on the employed cues are being classified in a recent study [11] where they are divided into three major categories.

The first category is a method to detect face liveness which relies on image quality/distortion measures. Work in [12] which consists of identifying print attacks using the difference in the 2D Fourier spectra is an example of the method in this category. The work stated in [13] utilizes the Lambertian model which comprises of variational Retinex based approach and Gaussian filters difference as its two methods. The work in [14] uses power spectrum and local binary patterns [15] to exploit both frequency and texture information. [16] modeled spatial and temporal information for face presentation attack. [17] proposes the combination of motion and texture methods via score level fusion. Difference-of-Gaussian filters to choose specific frequency bands for feature extraction was done in [18]. The work done in [19] proposes presentation attack detection by analyzing the texture represented using multi-scale local binary pattern [15] which provides a unique feature space for coupling spoofing detection and face recognition. The results from [20] reported good performance on Replay-Attack database.

The second category uses methods which are based on detecting different signs of vitality which make use of characteristics corresponding to live faces. For example, presentation attack detection in [21] uses blinking which is used with others cues in other work. Such as [22] recommend the use of all the dynamic information content of the video represented using dynamic mode decomposition method. The work done in [23] utilizes both eye-blink and scene content clues as a hybrid face liveliness detection system against spoofing with photographs, videos, and 3D models of a valid user in a face recognition system.

The last category consists of methods based on the difference in motion patterns between real and presentation attacks. It is assumed that the presentation attack have rigid motion whereas real-access attempts has both rigid and non-rigid motion. This approach depends on the fact that real accesses correlate with 3D structures whereas presentation attack media are often at 2D planes. Eulerian motion magnification using two sets of features composed of LBPs(Local Binary Patterns) [15] to enhance facial expressions is a typical case of the method in this category. The new liveness detection method is proposed in [24] which utilizes the difference in optical flow fields generated by the movements of 2D planes and 3D objects. A countermeasure against face presentation attack was proposed in [4] which were based on foreground/background motion correlation using optical flow showing promising results on the Photo-Attack database. The work in [3] used geometric invariants to detect replay attacks once a set of automatically located facial points are detected which was evaluated on two publicly available databases of NUAA [13] and HONDA [25].

While most of the existing methods use real access data to try and learn a general classifier to outline presentation attack attempts, work in [26] uses both texture and motion cues, the authors built two presentation attack detection methods, one being a generative approach while the other being a discriminative method to study the client-specific information embedded in the feature space and its effects on the performance of the system. Similarly, the work in [27] proposes a method using a classifier trained explicitly for each subject.

The current work regarding detection mechanism share some similarities to the existing approach which utilizes image content representation is distinct in the way we formulate the existing the detection problem. The standard approach used to detect an anomaly in an image uses two-class formulation where they separate the negative from the positive samples, our proposition uses one-class pattern classification methods, testing it in a modified as well as an existing method which yields good results to identify presentation attack attempts. Moreover, the evaluations are performed by using a custom-made database which better reflects the difficulties of detection in realistic scenarios. Also, many of the existing papers are supervised and conducted using face images/videos. These papers are evaluated using public databases such as Replay Attack Database. However, all of these public databases aims at the evaluation of counterfeit samples that imitate living organisms and does not assume the possibility of attacks that are performed by using unknown samples.For this reason, in this paper, we created our custom-made database of palm for evaluation by considering the possibility of various unknown samples. The reason for choosing Palm as a modality is that it is smaller in size as compared to face, the database can be made easily with an inexpensive camera, and it is easy to create an unknown counterfeit of the entire palm by wearing gloves or by making a false palm out of different compounds.

3 PRESENTATION ATTACK DETECTION USING ANOMALY GAN

3.1 Generative Adversarial Networks

Goodfellow et al. introduced a concept of Generative Adversarial Network(GAN) [28] which learns a *generator* expression indistinguishable by a *discriminator* by training a *generator* model and *discriminator* model simultaneously. The aim of the *generator* is to fool the *discriminator* by learning the probability distribution of the input samples. Let $\boldsymbol{x}$ be an input sample whose true probability distribution is $p(\boldsymbol{x})$. G is a *generator* that takes a latent vector $\boldsymbol{z}$ randomly selected from the latent space $\mathcal{Z}$ and outputs a new sample $G(\boldsymbol{z})$. The *discriminator* D then outputs the probability that the given input is either the true input from $p(\boldsymbol{x})$ or the $G(\boldsymbol{z})$ from the *generator*. These two models are simultaneously trained using the min-max game of the formula:

$$\min_D \max_G V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1-D(G(\boldsymbol{z})))] \tag{1}$$

Radford et al. [29] introduced deep convolutional generative adversarial networks (DCGAN) for unsupervised learning of features by utilizing convolutional neural networks as the *generator* and *discriminator* network. More specifically, they replaced the pooling layer with stride convolution layer so that the network can learn its own spatial upsampling. Additionally, they removed the full connection layer at the top of the convolution feature to improve model stability. Finally, batch normalization was utilized to suppress training problems caused by poor initialization and helps the propagation of gradients in deep models by normalizing each unit to have zero mean and unit variance.

3.2 Proposed Anomaly GAN for PAD

To detect presentation attack using a single image, we propose unsupervised learning to identify anomalies in imaging data as candidates for the fake sample. Fig. 1 shows an overview of our proposal. Our proposed scheme is based on unsupervised anomaly detection scheme proposed in [30] which is aimed at detection of disease markers in medical imaging(hereafter, AnoGAN). AnoGAN uses DCGAN to learn a manifold of live sample variability, accompanying an anomaly scoring scheme based on the mapping from image space to a latent space.

3.2.1 Palm Imaging Model

We learn the palm image manifold $\mathcal{X}$ on the image space with unsupervised learning using only the live palm images. When a query image is not included in the learned manifold $\mathcal{X}$, it can be detected as an unknown input.In DCGAN [29], *generator* uses latent vector $\boldsymbol{z}$ chosen from latent space $\mathcal{Z}$ uniformly at random to obtain a smooth mapping $G(\boldsymbol{z})$ to palm image manifold $\mathcal{X}$.

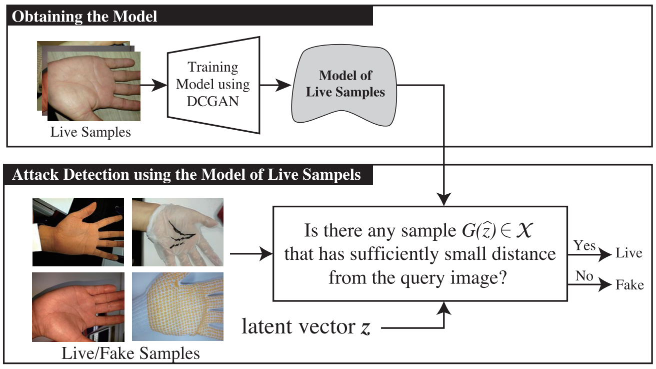

Obtaining the Model
Live Samples
Training Model using DCGAN
Model of Live Samples
Attack Detection using the Model of Live Sampels
Is there any sample $G(\hat{z}) \in \mathcal{X}$ that has sufficiently small distance from the query image?
Yes
Live
No
Fake
latent vector z
Live/Fake Samples

Figure 1: Overview of our proposal.

3.2.2 Deriving Latent Vector

We can detect the Presentation Attack by checking whether query image $\boldsymbol{x}_q$ is included in the palm image manifold $\mathcal{X}$ learned in the clause 3.2.1. Since DCGAN calculates $G(\boldsymbol{z})$ using the randomly chosen latent vector $\boldsymbol{z}$, $G(\boldsymbol{z})$ corresponds to a random point on the palm image manifold $\mathcal{X}$. Consequently, the distance between $G(\boldsymbol{z})$ and the query image $\boldsymbol{x}_q$ does not necessarily become small even if it is a live sample. Therefore, to detect an anomaly sample, we should confirm the existence of a latent vector $\widehat{\boldsymbol{z}}$ that has a sufficiently small distance between the query image $\boldsymbol{x}_q$ and $G(\widehat{\boldsymbol{z}})$ on the manifold $\mathcal{X}$.

For finding the $\widehat{\boldsymbol{z}}$ from randomly chosen latent vector $\boldsymbol{z}$, we use the backpropagation approach proposed in [30]. The loss function $\mathcal{L}(\boldsymbol{z}_\gamma)$ for backpropagation is defined as follows:

$$\mathcal{L}(\boldsymbol{z}_\gamma) = (1-\lambda)\cdot\mathcal{L}_R(\boldsymbol{z}_\gamma) + \lambda\cdot\mathcal{L}_D(\boldsymbol{z}_\gamma) \tag{2}$$

where $\boldsymbol{z}_\gamma$ is an updated latent vector to fool *discriminator* D, $\mathcal{L}_R(\boldsymbol{z}_\gamma)$ is the generator loss, $\mathcal{L}_D(\boldsymbol{z}_\gamma)$ is the discriminator loss and λ is a fixed parameter for convex combination. The residual loss and the discriminator loss can be obtained as follows:

$$\mathcal{L}_R(\boldsymbol{z}_\gamma) = \sum |\boldsymbol{x}_q - G(\boldsymbol{z}_\gamma)| \tag{3}$$

$$\mathcal{L}_D(\boldsymbol{z}_\gamma) = \sum |\boldsymbol{f}(\boldsymbol{x}_q) - \boldsymbol{f}(G(\boldsymbol{z}_\gamma))| \tag{4}$$

where $\boldsymbol{f}(\cdot)$ is an output of the *discriminator* function. Only the coefficients of $\boldsymbol{z}$ are adapted via backpropagation. The trained parameters of the *generator* model and *discriminator* model are kept fixed. In our proposal, $\widehat{\boldsymbol{z}}$ is obtained by applying backpropagation process α times with query image $\boldsymbol{x}_q$ and randomly selected $\boldsymbol{z}$. The obtained $\widehat{\boldsymbol{z}}$ is used in classification process.

3.2.3 Classification

In classification process, we investigated the three types of score function, anomaly score $A(\boldsymbol{x})$, residual score $R(\boldsymbol{x})$, and discriminator score $D(\boldsymbol{x})$, respectively. The relationship between each score is defined as follows:

$$A(\boldsymbol{x}_q) = (1-\lambda)\cdot R(\boldsymbol{x}_q) + \lambda \cdot D(\boldsymbol{x}_q) \tag{5}$$

where the residual score $R(\boldsymbol{x}_q)$ and discrimination score $D(\boldsymbol{x}_q)$ are defined by the residual loss $\mathcal{L}_R(\widehat{\boldsymbol{z}})$ and discriminator loss $\mathcal{L}_D(\widehat{\boldsymbol{z}})$ using at the α update iteration of the mapping procedure to the latent space, respectively. All score functions output a large score for an anomaly image. In our experiments, we use λ = 0.9 in equations (2) and (5) which was found empirically due to preceding experiments on our palm dataset.

3.2.4 Cumulative score calculation

The experiment performed in [30] requires the trained model to be executed for α times and also requires to perform numerous backpropagation steps even if the sample was obviously an anomaly sample. Therefore, We propose to utilize a cumulative score $C(x_q, \beta)$ for a query image x_q at β-th backpropagation step which is as follows:

$$C(x_q, \beta) = \sum_{b=0}^{\beta} A(x_{q,b}) \tag{6}$$

where $A(x_{q,b})$ is an anomaly score for b-th backpropagation step.

If the target is a live sample, $A(x_{q,b})$ will decrease more sharply as the backpropagation step increases since $G(z)$ and live sample are within the same manifold $\mathcal{X}$. Therefore, if we assume that α is the maximum count for the execution of backpropagation, then the input sample can be classified as a live sample if the value of cumulative score $C(x_q, \alpha)$ is smaller than the threshold th. On the other hand, if at a certain point β whose value is $\beta < \alpha$, the calculation can be canceled and the input sample can be classified as a fake image as soon as the cumulative score satisfies $C(x_q, \beta) > th$. For this reason, it is possible to reduce the amount of calculation as compared with the usual method which always requires α times calculation for the backpropagation.

4 EXPERIMENT

In this section, first, a description of the custom-made database and the evaluation protocols used in this experiment is provided, following by experimental results obtained from the database used. All the experiments were carried out using Python with the tensorflow and pytorch library on a machine with configuration (Intel i7-5930K, 64GB RAM, 12x Intel(R) Core(TM), Ubuntu 64bit) environment.

4.1 Database

Many previous works have used public live/fake dataset such as Replay-Attack Database [2] and Unconstrained Smartphone Spoof Attack (USSA) Database [31]. However, it contains only a specific type of fake photo and video samples making it inadequate in terms of anomaly samples. Therefore, in our experiment, we constructed a custom-made database to make sure that the system is being able to make a clear distinction between live and fake samples even when the system encounters unexpected inputs such as palm with a glove, palm with a vinyl glove, etc. which have no direct relation with the hand. So, in our custom-made database, we prepared a large amount of data to check whether the system will be able to counter any fake sample provided to it by the attacker as an input.

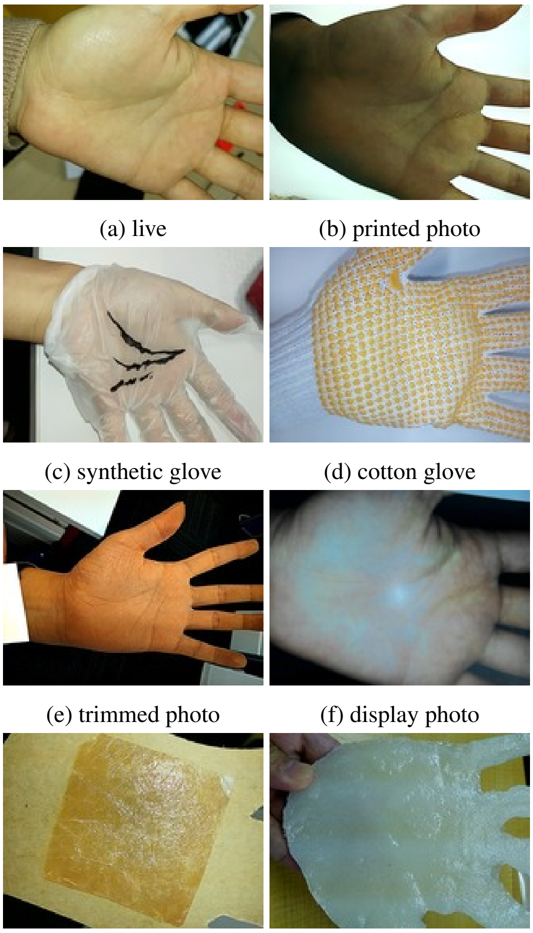
The custom-made database used in the experiment consists of 8748 live samples and 6648 fake samples of palm with an image resolution of 160x120 pixels taken directly from approximately 2000 people with ten different types of mobile cameras (LG G5, LG Nexus 5x, LG Nexus 5, Sony Xperia X Performance, Elephone P9000, Sharp Aquos SHV34, Doogee X5max, Huawei GR5 (KII-L22), ASUS zenfone2 (Z00D), ASUS P008). The images are taken in different non-controlled indoor surrounding conditions such as, inside office with different background or inside the building with varying conditions of lighting which also includes photo that is made in a dark place with the help of flashlight ,etc with varying postures in order to anticipate all kind of possibilities of the images that will be used as the input for the system. The training set used to train the AnoGAN model comprises of randomly selected 8000 live samples. The test set in total consist of 7396 samples out of which 748 were live palm samples and 6648 were fake palm samples from cases not included in the training set. The training that we are performing in this experiment is uncontrolled without of external interference. Example of true samples and different variety of fake samples that were used while training the system is given as below in Fig. 2. In order to include as many variety of unexpected fake samples as possible to check the accuracy of the system, we included photos such as (b)printed photo, (c)hand wearing synthetic glove, (d)hand wearing cotton glove,(e)printed photo that were cut from the border of the hand part in order to resemble hand in 2D, (h)gelatin or (g)ham which may not have direct relation with the hand but can resemble skin and (f)photo that were taken from a digital device such as iPad or webcam.

4.2 Evaluation Protocol

The manifold of live images was solely learned on image data of live cases with the aim to model the variety of live appearance. So for that purpose, 8000 live samples are selected from the database as noted earlier to develop the model. In a real case scenario, it is difficult for users to collect 8000 images in order to create a model for such evaluation, but this problem can be solved by using learned models provided by vendors who have easy access to a large amount of data which will be used to generate the required model for the experiment, real-life use, etc. For performance evaluation in anomaly detection, we ran various protocols exploited by researchers.

4.2.1 Our Protocol

All the training and test conducted for the anomaly detection in this work are based on the one class system where only the



(a) live (b) printed photo

(c) synthetic glove (d) cotton glove

(e) trimmed photo (f) display photo

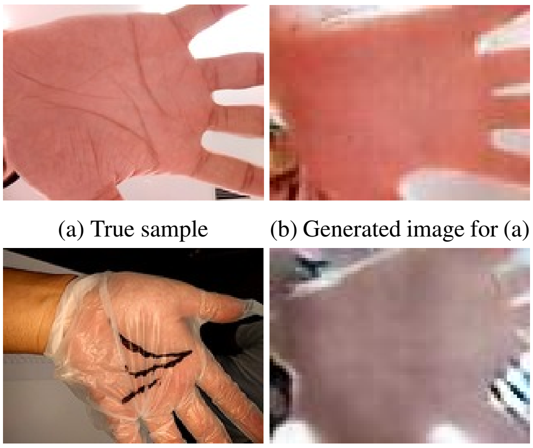
(g) ham (h) synthetic compound

Figure 2: Example samples used for training the model for 1 class system. (a) is an example of true sample used for training the model and (b) to (h) are the different variety of fake samples that were used for testing the model.

live samples are used to develop the model. In particular, the following systems are used for the development and evaluation:

- AnoGAN+RAW: The AnoGAN which uses one class system trained using the original image
- SVM1+RAW: The one-class SVM with a Gaussian kernel trained using the original image
- SVM1+LBP: The one-class SVM with a Gaussian kernel trained using the LBP feature

For each of these protocols, the model was trained using 8000 live samples and the test was conducted by using 100 fake samples along with 100 live samples which were not included in 8000 live samples that were used for training the model to check the accuracy of the model in order to distinguish the fake samples from the live samples. Residual score $R(\boldsymbol{x}_q)$ is taken into account in order to differentiate between live samples and fake samples. The purpose of this unsupervised one class training is to find the epochs whose training accuracy as well as prediction accuracy are good and which does not cause overfitting. For this dataset the epochs which showed the best result is 50. As we try to increase the epochs the accuracy of the model decreased. In this experiment, the unsupervised learning was conducted by changing the epochs as 20, 25, 50,$\cdots$,100. In the result section of this experiment, we used the result of 25 epochs as the representative example and the result of 50 epochs as it shows the best result. The residual score can vary each time the test is conducted even if the image used for testing is the same because the residual score measures the visual dissimilarity between query image $\boldsymbol{x}_q$ and generated image $G(\widehat{\boldsymbol{z}})$ in the image space by finding a point $\widehat{\boldsymbol{z}}$ in the latent space that corresponds to an image $G(\widehat{\boldsymbol{z}})$ that is visually most similar to query image $\boldsymbol{x}_q$ and that is located on the manifold $\mathcal{X}$. We ran 100 backpropagation steps (α = 100) for the mapping of new images to the latent space $\mathcal{Z}$. The image produced after performing 100 backpropagation is given as in Fig. 3 which shows that the trained model can generate reasonably realistic looking images when a live sample is used for the classification of the image as the image is generated from inside the manifold $\mathcal{X}$. On the other hand, when a fake sample is used for the classification process, as the images are obtained from the same manifold, images that are close to real to some extent can be obtained. However, as the query sample is unreal, the residual score gets bigger.



(a) True sample (b) Generated image for (a)

(c) Fake sample (d) Generated image for (c)

Figure 3: Example samples used for training the model for 1 class system and the respective image produced by the AnoGAN after performing 100 backpropagation. (a)True sample (b)Image generated by AnoGAN for true sample (c)Fake sample (d)Image generated by AnoGAN for fake sample.

4.2.2 Evaluation Metric

For evaluating the result obtained, we consider the Area Under Curve (AUC) obtained from Receiver Operating Characteristic (ROC) curves. The ROC curve was made using the residual score as the parameter which yields good results as shown in [30]. The vertical axis and the horizontal axis of ROC curves usually present True Positive and False Positive Rate respectively. It indicates that the plot ’s top left corner is the optimal point. Preferable TPR for the ROC curve is equal to one which makes the excellent AUC’s values approaching one.

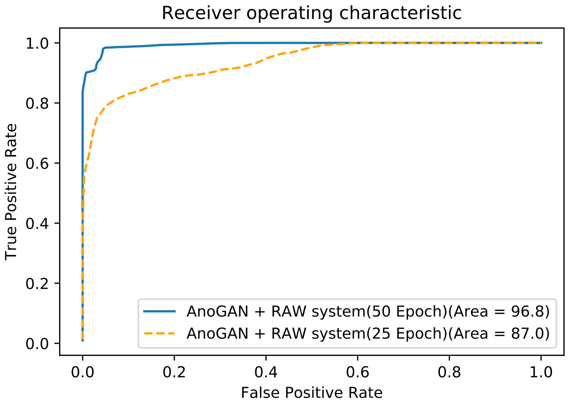


Figure 4: The above figure represents the ROC graph of the AnoGAN model trained for 25 (orange) and 50 (blue) epochs respectively by using live samples as an input image for training.

Table 1: Area under the ROC (AUC) (%) for different systems obtained by using custom database.

System	AUC(%)
AnoGAN+RAW (20 Epoch)	83.1
AnoGAN+RAW (25 Epoch)	87.0
AnoGAN+RAW (50 Epoch)	96.8
SVM1+RAW	34.3
SVM1+LBP	83.5
SVM1+LBP (cos similarity)	75.9

4.3 Evaluation Results

The one-class systems introduced earlier are evaluated on the custom-made database which used 8000 live samples to develop the model. To make sure that there is no bias in the result obtained after testing each of the models we took out a total of 200 samples randomly from the palm database, 100 samples each from live samples and fake samples. For the fake samples, even though the 100 images taken out were selected at random, it was made sure that it contained all the variety of samples that were taken into account while creating the fake samples. By doing so, we can see to what extent the trained model produces the desired result even if it encounters unexpected input which is fake but has no direct relation with hand.

Figure. 4 represents the ROC graph of the results obtained from different models where the Y-axis shows the True Positive Rate and the X-axis shows False Positive Rate. Additionally, Table 1 and 2 show the AUC and HTER (Half Total Error Rate) respectively. Note that HTER can be calculated by $\min(TN + FP)/2$.

4.3.1 Accuracy

Table 1 shows that the best performing one-class system regarding average performance is Ano-Gan+RAW with an average AUC of 96.8%. The result obtained regarding AUC by using one class SVM system [32] as a model for training and testing the dataset as used for training and testing AnoGAN is also shown in Table 1 and 2. It is clearly visible that the proposed AnoGAN system is far more better than the conventional one class SVM system. As far as the security check for AnoGAN is concerned, it can be conducted by using the image produced by the AnoGAN as the input image while testing the model and comparing the residual score with that of the unseen real samples and fake samples. As far as the one class SVM system are concerned, SVM1+LBP performed better as compared to SVM1+RAW. SVM1+LBP is more sensitive whereas SVM1+RAW is less sensitive to the attack model because as stated in [19] by using LBP feature they were able to perform their experiment in a robust way which was computationally fast and didn't required any user-cooperation. Moreover, the extensive experimental analysis done by them on a publicly available database showed excellent results compared to existing works which proves clarifies that SVM1+LBP will show better results as compared to SVM1+RAW.

Table 2: Half Total Error Rate(HTER)(%) for different sys-tems obtained by using custom database.

System	HTER(%)
AnoGAN+RAW (20 Epoch)	17
AnoGAN+RAW (25 Epoch)	13
AnoGAN+RAW (50 Epoch)	3
SVM1+RAW	34
SVM1+LBP	17
SVM1+LBP (cos similarity)	20

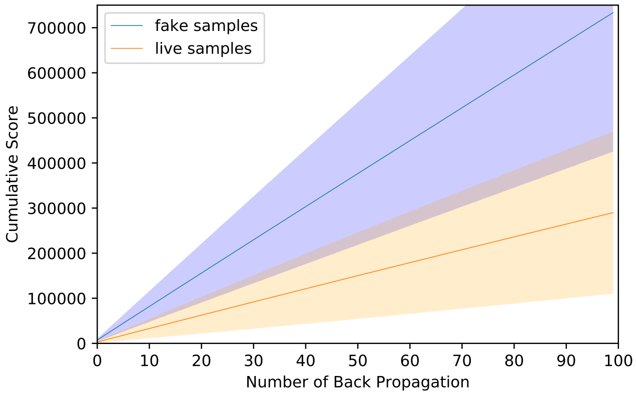


Figure 5: The above figure represents cumulative score calculation of the AnoGAN model trained for 50 epoch where the average of cumulative score for each epoch along with the standard deviation is taken into consideration.

4.3.2 Computational efficiency

As described in section 3.2.4, when using the cumulative score to perform the analysis, fake input can be detected at β backpropagation which is less than the maximum number of backpropagation α. At this time, in order to calculate the efficiency we use the ratio of the average number of backpropagation $\overline{\beta}$ and the maximum number of backpropagation α that

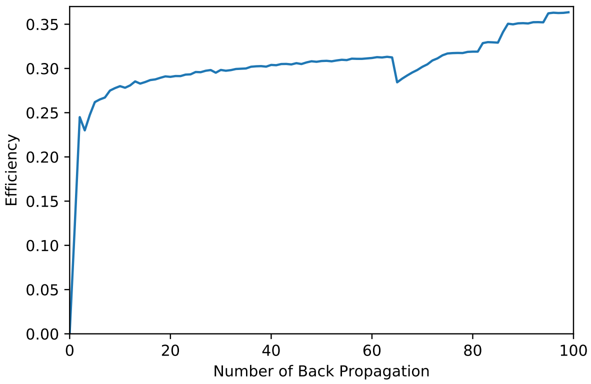


Figure 6: The above figure represents the possible efficiency of the AnoGAN model on the basis cumulative score at each backpropagation.

is $\overline{\beta}/\alpha$ as efficiency and evaluate the computational effectiveness of cumulative score.

Figure. 6 shows the value of efficiency when α is changed. The protocol mentioned in section 4.2 was used for training and evaluation of AnoGAN. Regarding the threshold th for each α, we calculated and applied the threshold value that minimizes HTER in test data. As it can be seen from Fig. 5, the larger the α, the greater the efficiency and the greater the effect of the cumulative score. Also, since efficiency > 0 is always valid except for $\alpha = 0$, it is confirmed that using cumulative score always has a computational advantage compared to using residual score.

5 DISCISSION

In the one-class system, the AnoGAN method is more accurate as compared to conventional one-class SVM which produced good results in other researches. Among the models that were produced by the AnoGAN system, the model which was trained for 50 epoch showed the better result as compared to other models. From the results we obtained by performing this, we can see that an unsupervised model such as AnoGAN could have many benefits, we also see some research limitations.

First, the number of epochs for which you have to train the system may depend on the number of images that you are using to train the system. However, we have not yet found out the relation between them. Therefore, finding the relationship and optimizing the number of epoch needed from training the model could help us improve our accuracy.

Second, Infinite samples can be produced from the DCGAN used in AnoGAN as it digitally produces images which are considered as a real sample by the system generating the model which can be used to improve the accuracy of the system and give better results while calculating the residual score for a given input. Here a doubt arises that, as the DCGAN that we are using while performing this experiment can generate an infinite number of samples, it can be thought that the attacker can misuse the produced image by using it to attack the system, but it is unlikely to happen. This is because, even when the result of the GAN is obtained as an image, it is necessary to output it into some other form to make it visible to us such as paper or a display and shoot with a camera. In this method, images of living organisms are produced onto a paper or a display, and samples that are photographed by cameras are detected as a fake sample with high probability. So, it is not possible to misuse the output of the DCGAN. On the other hand, as features of biometric information such as the face, palm, etc. differ depending on the modality, it was considered possible to perform the estimation of bio-distribution from different modality using DCGAN. However, it was not fully verified as we didn't had many unknown sample in our custom-made database. It is necessary to do the future experiment by including more unknown sample in the data set.

Third, While testing the one-class SVM method the SVM1 + LBP produced better results as compared to SVM1 + RAW. Therefore if the code of AnoGAN is designed in such a way that it calculates the loss function while taking LBP (histogram, cosine similarity) into consideration from the point of training, then there is a chance that it might produce a better result. Also, we have only examined anomaly detection systems based on 1 class SVM and AnoGAN; it would be better if we study other anomaly detection approach also.

Fourth, it can be concluded that even if you train the system by using the true samples only, it does not perform well enough and more modification and research should be conducted to improve the performance of this type of system. Therefore, to improve the performance, we would like to take the cost included in the making of the data set which would serve as a checkpoint from where we can strive for further improvement.

Last, in the experiment conducted this time, we have only used the image captured by using the camera that can be found in any of the typical mobile used by us in our day to day life. We also did not control or limit the posture of the palm at the time of the shooting. This is done to increase the robustness of the model as we assume that by doing so the system will be able to adjust its identification analysis concerning the user's natural behavior. However, since as we are using palm + mobile in this experiment, the number of possible postures will get limited to some extent and problems may arise when this system is applied to entirely different applications.

In the future experiment, it is possible that we can incorporate a higher degree of living body detection which can distinguish between a live and a fake sample with even more accuracy by combining the system used for creating the model with sensors that measure biological reaction such as ECG.

6 CONCLUSION

In this study, we investigated a palm presentation attack detection method based on an anomaly detection using Generative Adversarial Network. Our remarkable result is that the proposed PAD scheme achieved 96.8% AUC and 3% of HTER by using a model trained only with real samples. It is visible that our proposal can achieve a far better result than conventional one-class SVM systems. Additionally, it should be noted that our method can detect presentation attack by using a single static image. Therefore, this method can also be directly applied to deal with video presentation attack or be integrated with a video-based palm liveness detection method for better performance. It is left to investigate about loss function suitable for presentation attack and reduce the number of backpropagation α to improve our method more secure and convenient.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Number JP18K11294. We thank Eizaburo Iwata and Hiroki Kamanaka for useful discussions. The Palm Presentaton Attack Dataset is provided by Normee co, ltd.

REFERENCES

[1] A Pacut and A Czajka. "Aliveness Detection for IRIS Biometrics,". In *Carnahan Conferences Security Technology, Proceedings 2006 40th Annual IEEE International*, pages 122–129. IEEE, (2006).

[2] A. Anjos and S. Marcel. "Counter-measures to photo attacks in face recognition - A public database and a baseline,". In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7. IEEE, (2011).

[3] M. De Marsico, M. Nappi, D. Riccio, and J. Dugelay. "Moving face spoofing detection via 3D projective invariants,". In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 73–78. IEEE, (2012).

[4] A. Anjos, M. M. Chakka, and S. Marcel. "Motion-based counter-measures to photo attacks in face recognition,". *IET biometrics*, 3(3):147–158, (2013).

[5] H. Choi, R. Kang, and J. Choi, K.and Kim. "Aliveness Detection of Fingerprints using Multiple Static Features,". In *Proc. of World Academy of Science, Engineering and Technology*, pages 201–205, (2007).

[6] R. N. Rodrigues, N. Kamat, and V. Govindaraju. "Evaluation of biometric spoofing in a multimodal system,". In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–5, (2010).

[7] G. Fumera G. L. Marcialis F. Roli B. Biggio, Z. Akthar. "Robustness of multi-modal biometric verification systems under realistic spoofing attacks,". In *Biometric Measurements and Systems for Security and Medical Applications (BIOMS) 2011 IEEE Workshop on*, pages 1–6. IEEE, (2011).

[8] P. Wild, P. Radu, L. Chen, and J. Ferryman. "Towards anomaly detection for increased security in multibiometric systems: Spoofing-resistant 1-median fusion eliminating outliers,". In *IEEE International Joint Conference on Biometrics*, pages 1–6. IEEE, (2014).

[9] T. Ohki and A. Otsuka. "Theoretical vulnerabilities in map speaker adaptation,". In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2042–2046. IEEE, (2017).

[10] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu. "DolphinAttack: Inaudible Voice Commands,". In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, pages 103–117. ACM, (2017).

[11] T. de Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikäinen, and S. Marcel. "Face liveness detection using dynamic texture,". *EURASIP Journal on Image and Video Processing*, 2014(1):2, (2014).

[12] J. Li, T. Wang, Y.and Tan, and A. K. Jain. "Live face detection based on the analysis of fourier spectra,". In *Biometric Technology for Human Identification*, volume 5404, pages 296–304. International Society for Optics and Photonics, (2004).

[13] X. Tan, Y. Li, J. Liu, and L. Jiang. "face liveness detection from a single image with sparse low rank bilinear discriminative model,". In *European Conference on Computer Vision*, pages 504–517. Springer, (2010).

[14] G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim. "Face liveness detection based on texture and frequency analyses,". In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 67–72. IEEE, (2012).

[15] T. Ojala and D. Harwood M. Pietikainen. "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,". In *Image Processing. Proceedings of the 12th IAPR International Conference on*, pages vol. 1, pp. 582–585. IEEE, (1994).

[16] W. R. Schwartz, A. Rocha, and H. Pedrini. "face spoofing detection through partial least squares and low-level descriptors,". In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, (2011).

[17] J. Komulainen, A. Hadid, M. Pietikäinen, A. Anjos, and S. Marcel. "Complementary countermeasures for detecting scenic face spoofing attacks,". In *Biometrics (ICB), 2013 International Conference on*, pages 1–7. IEEE, (2013).

[18] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. "A face antispoofing database with diverse attacks,". In *Biometrics (ICB), 2012 5th IAPR international conference on*, pages 26–31. IEEE, (2012).

[19] J. Määttä, A. Hadid, and M. Pietikäinen. "Face spoofing detection from single images using micro-texture analysis,". In *Biometrics (IJCB), 2011 international joint conference on*, pages 1–7. IEEE, (2011).

[20] I. Chingovska, A. Anjos, and S. Marcel. "on the effectiveness of local binary patterns in face anti-spoofing,". In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, pages 1–7. IEEE, (2012).

[21] G. Pan, L. Sun, Z. Wu, and S. Lao. "Eyeblink-based anti-spoofing in face recognition from a generic webcamera,". In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, (2007).

[22] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. TS Ho. "Detection of face spoofing using visual dynamics,". *IEEE transactions on information forensics and security*, 10(4):762–777, (2015).

[23] G. Pan, L. Sun, Z. Wu, and Y. Wang. "Monocular

camera-based face liveness detection by combining eye-blink and scene context,”. *Telecommunication Systems*, 47(3-4):215–225, (2011).

[24] W. Bao, H. Li, N. Li, and W. Jiang. “A liveness detection method for face recognition based on optical flow field,”. In *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on*, pages 233–236. IEEE, (2009).

[25] K. Lee, J. Ho, M. Yang, and K. “Visual tracking and recognition using probabilistic appearance manifolds,”. *Computer Vision and Image Understanding*, 99(3):303–331, (2005).

[26] I. Chingovska and A. R. D. Anjos. “On the use of client identity information for face antispoofing,”. *IEEE Transactions on Information Forensics and Security*, 10(4):787–796, (2015).

[27] J. Yang, Z. Lei, and S. Z. Yi, D.and Li. “Person-specific face antispoofing with subject domain adaptation,”. *IEEE Transactions on Information Forensics and Security*, 10(4):797–809, (2015).

[28] J. I. Goodfellow, 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).

[29] A. Radford, L. Metz, and S. Chintala. “Unsupervised representation learning with deep convolutional generative adversarial networks,”. *arXiv preprint arXiv:1511.06434*, (2015).

[30] T. Schlegl, P. Seeböck, Sebastian M. Waldstein, U. Schmidt-Erfurth, and G. Langs. “Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery,”. In *Proceedings of the 25th International Conference of the Information Processing in Medical Imaging IPMI 2017, Boone, NC, USA*, pages 146–157, June (2017).

[31] K. Patel, H. Han, and A.K. Jain. “Secure Face Unlock: Spoof Detection on Smartphones,”. In *IEEE Trans. Information Forensic and Security*, June (2016).

[32] A. Smola J. Shawe-Taylor J. Platt B. Schölkopf, R. Williamson. “Support vector method for novelty detection,”. In *Proceedings of the 12th International Conference on Neural Information Processing Systems*, pages 582–588, (1999).

(Received October 20, 2018)
(Revised December 28, 2018)

Vishu Gupta has completed his high school from Sri Venkateshwas International School, New Delhi, India. He is currently a bachelor student within the computer science program of Faculty of Informatics, Shizuoka University. He will graduate with a B.Tech in Computer Science in 2019. His research interests include machine learning algorithms, pattern recognition and their security.

Masakatsu Nishigaki has received his Ph.D. in Engineering from Shizuoka University, Japan. He served as a Postdoctoral Research Fellow of the Japan Society for the Promotion of Science in 1995. Since 1996 he has been engaged in research at the Faculty of Informatics, Shizuoka University. He is now a Professor at the Graduate School of Science and Technology of Shizuoka University. His research interests are in wide variety of information security, especially in humanics security, media security, and network security. He served as Chief Examiner of IPSJ (Information Processing Society of Japan) Special Interest Group on Computer Security from 2013 to 2014, Chair of IEICE (Institute of Electronics, Information and Communication Engineers) Technical Committee on Biometrics from 2015 to 2016, and currently serving as Director of JSSM (Japan Society of Security Management) since 2016. He is IPSJ (Information Processing Society of Japan) fellow.

Tetsushi Ohki has received the BE and ME degrees in electronics and communication engineering from Waseda University, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in Engineering from Waseda University in 2010. He is currently a Lecturer at the Graduate School of Science and Technology of Shizuoka University, Japan. His research interests include biometrics, pattern recognition, information security and privacy. He is a member of IEICE (Institute of Electronics, Information and Communication Engineers) and IPSJ (Information Processing Society of Japan) .