Towards Fingerprint Presentation Attack Generation using Generative Adversarial Networks

Leonardo Capozzi[1,2]
leonardo.g.capozzi@inesctec.pt
Tiago Gonçalves[1,2]
tiago.f.goncalves@inesctec.pt
Jaime S. Cardoso[1,2]
jaime.s.cardoso@inesctec.pt
Ana Rebelo[3]
ana.maria.s.rebelo@accenture.com

[1] Faculdade de Engenharia
Universidade do Porto
Porto, Portugal
[2] INESC TEC
Porto, Portugal
[3] Accenture Portugal
Lisboa, Portugal

Abstract

Most of the available systems rely on fingerprint recognition and have shown to be reliable in terms of accuracy, speed and purported security. However, they also present several vulnerabilities against spoof attacks. To overcome this flaw, several automated spoofing detection models have been developed, but they end up assuming that spoof detection is a binary closed-set problem, which is not realistic. Recent works have proposed the use of adversarial methodologies to improve the model's generalisation capacity to unseen spoof attacks. Following this research line, we performed a study on the application of generative adversarial neural networks (GANs) for the generation of synthetic data to be employed during the model's training. We hypothesise that by using GANs, one could learn a distribution that could contain all possible spoofing attacks, thus opening the possibility to learn classifiers that could be more robust. In this work, we optimised a GAN and conditional GAN (cGAN) to generate synthetic images of real and fake fingerprints and used this data in the training of classifiers for the detection of single spoofing attacks.

1 Introduction

Several studies revealed that recognition systems based on fingerprint recognition are highly vulnerable to spoof attacks (*e.g.*, a 2D printing with conductive ink to replicate the fingerprint of a victim left behind on their keyboard) [1]. Therefore, intending to overcome this flaw, several automated spoofing detection systems have been developed, to detect and flag spoof attacks before performing biometric authentication [1]. The main drawback about these systems is that they assume that spoof detection is a binary closed-set problem, *i.e.*, *live* or *single-material spoof* [1]. For this reason, most of the methods may be overestimated regarding their performance metrics, since they end up using only one type of spoof samples to train and test the models [6]. If we take into account the diversity of materials (with different mechanical and optical properties) that can be employed to produce different spoofs, we may conclude that this may be, in fact, an open-set problem, where we should assume *a priori* that we do not know all the existing spoofs [1]. The problem of presentation attack detection (PAD) generalisation is not new. Besides, [9] stated that every time we present a new presentation attack (PA) species in the test phase, the performance of the classifier drops. Several methodologies based on one-class classification for the *bona fide* (BF) class or adversarial training [1] have been proposed to increase the robustness of the classifier. However, the current paradigm has not changed: most of the recent proposals either rely on binary classification approaches or use part of the data available at training time to design the models. These two possibilities still make very optimistic assumptions regarding the attacker [5]. Following the work proposed by Pereira et al. [5], we intend to extend the adversarial methodology and use generative adversarial networks (GANs) to 1) increase the quality of the generated PA species; and 2) increase the robustness of the PAD method against unseen attacks. The intuition behind this approach is supported by the following hypothesis: if one could learn a distribution that contains all possible PAs, then it should be possible to learn a function that could correctly discriminate against BF and unseen PAs, without the need of explicitly using them during the training phase. The code related to this work is available in a GitHub repository [1].

[1] https://github.com/leonardogomesc/Image-to-image-Generative-Adversarial-Network

2 Related Work

Generative Adversarial Networks Generative Adversarial Networks (GANs) [2] have become very popular for image generation, and some variations to the original methodology have been proposed. It consists of two networks, a generator, and a discriminator. The goal of a GAN is to train a generator that creates images that are indistinguishable from images in the train set. The generator is constantly trying to generate images that "fool" the discriminator into thinking they are real, and the discriminator is constantly trying to distinguish between real images and generated images.The loss function of a GAN can be written as:

$$\mathcal{L} = \mathbb{E}_x[\log D(x)] + \mathbb{E}_y[\log(1 - D(G(y)))] \tag{1}$$

where G is the generator, D is the discriminator, x is a real image and y is a noise vector. The generator is optimised to minimise this loss function, while the discriminator is optimised to maximise it.

Conditional Generative Adversarial Network Since the input of the network is a random vector we have limited control over the generated images, which would be very useful for some problems. Conditional Generative Adversarial Networks (cGANS) [4] aim to solve this problem by giving the generator and the discriminator an additional input, a label that allows us to condition the generator to output an image belonging to a certain class or having a specific attribute present. The noise vector is still used to allow for some variation in the generated images belonging to the same class. The loss function of a cGAN can be written as:

$$\mathcal{L}_c = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x,z)))] \tag{2}$$

where x is a class label, y is a "real" image from class x and z is a random noise vector. The generator (G) tries to minimise this loss, and the discriminator (D) tries to maximise it. As training progresses the generator will output increasingly realistic images until they are practically indistinguishable from the images belonging to the train set.

3 Methodology

Data We used the LivDet2015 dataset, which was developed for the 2015 edition of the Liveness Detection Competition. This dataset is composed of five sub-datasets: Cross Match, Digital Persona, Green Bit, Hi Scan, and Time Series [6]. Our data processing pipeline is employed as follows: 1) images are converted into the grey-scale format; 2) images are then re-scaled into 110% of their original size through the addition of padding, with pixel values $= 255$; 3) a centre-crop operation is then applied to assure that images keep their final size as 64×64 and that the fingerprint is in the centre of the image as well; 4) the pixel values of the images are then re-scaled to be in the closed interval of $[-1, 1]$.

GAN and cGAN The GAN model we use in this work is based on the class of Deep Convolutional GANs (DCGANs) [7], and the cGAN model uses the U-Net architecture for the generator [8]. We start the training of the GAN with the generation of a latent-space vector nz of size 64 filled with random numbers sampled from a normal distribution with mean 0 and variance 1, *i.e.*, $nz \sim \mathcal{N}(0,1)$. This vector is given as input for the generator $G(\cdot)$, which will generate a synthetic fingerprint image $G(nz)$. We then feed the discriminator $D(\cdot)$ with $G(nz)$ to obtain its prediction $D(G(nz))$. The training of the cGAN starts with the generation of a noise

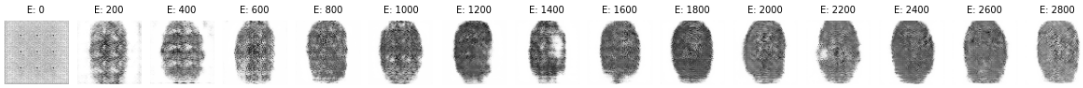

Figure 1: Examples of images generated by our GAN model per epoch (real fingerprints from the Cross Match dataset).

Table 1: Accuracy results obtained for the M1–M3 models for the different datasets and related materials. Best results highlighted in bold.

Material / Dataset	Cross Match			Digital Persona			Hi Scan		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
Ecoflex	**0.9130**	0.8859	0.8232	-	-	-	-	-	-
Playdoh	**0.9568**	0.8787	0.9062	-	-	-	-	-	-
Latex	-	-	-	**0.9648**	0.9640	0.9496	**0.9752**	0.9744	0.9712
Gelatine	-	-	-	0.9024	0.9112	**0.9160**	**0.9144**	0.9008	0.9064
Wood Glue	-	-	-	0.9384	**0.9392**	0.9384	0.8600	0.8704	**0.8776**

vector nz with a dimension of $64 \times 64 \times 1$, the same dimensions as the input image x (*bona fide* fingerprint). The noise vector is filled with random numbers sampled from a normal distribution with mean 0 and variance 1, *i.e.*, $nz \sim \mathcal{N}(0,1)$. Then it is concatenated to the input image in the channels dimension and given to the generator $G(\cdot)$, which will generate a synthetic fingerprint image $G(x,nz)$. We feed the discriminator $D(\cdot)$ with x and $G(x,nz)$ to obtain its prediction $D(x,G(x,nz))$. The discriminator is trained to classify the image y in $D(x,y)$ as being real or generated.

Fingerprint Presentation Attack Detection For the fingerprint presentation attack detection (FPAD) task, we used DenseNet121 [3] as backbone. We added a single neuron (linear layer) with the sigmoid activation at the end of the architecture as the classifier. Regarding data augmentation, we employ three different strategies: 1) Baseline training without data augmentation (M1); 2) Training with synthetic images generated with the GAN model, in which we add new images (*bona fide* and PA) generated with the different GAN models to the train set (M2); 3) Training with synthetic images generated with the cGAN model, in which we add new images (PA only) generated with the cGAN models to the train set (M3).

4 Results and Discussion

GAN and cGAN Figure 1 shows example images generated by our GAN model in different training epochs. When the GAN reaches equilibrium, we can observe that the synthetic image is now more realistic, whether it is from the *bona fide* or the PA classes. The cGAN models shows similar behaviour. Nevertheless, the process of generating new samples helped the training of the FPAD classifier in some cases.

FPAD Results Table 1 presents the accuracy results for the M1–M3 models. Regarding the performance of the FPAD models with or without data augmentation strategies based on a GAN or a cGAN, it is still not possible to conclude whether or not these strategies improve the model's predictive performance, however, one may observe that there is no clear evidence that their performance gets worse. Hence, this may suggest that the synthetic data we add during training may be part of the same distribution as the original data.

5 Conclusions and Future Work

Current results suggest that the predictive performance of the classifier is not jeopardised, which means that the synthetic data may be part of the same distribution as the original data. It is important to note that these experiments do not take into account unseen attacks (*i.e.*, models were just tested with the PAs that they were trained with). Future work should be devoted to improving the architectures of both the GAN and cGAN in the sense that they could be trained with higher resolutions.

Acknowledgements

This work is co-financed by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project NewSpacePortugal, with reference 11, and by FCT – Fundação para a Ciência e a Tecnologia within the PhD grants "2020.06434.BD" and "2021.06945.BD".

References

[1] Joshua J Engelsma and Anil K Jain. Generalizing fingerprint spoof detector: Learning a one-class classifier. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019.

[2] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[3] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[4] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014. URL `http://arxiv.org/abs/1411.1784`.

[5] Joao Afonso Pereira, Ana F Sequeira, Diogo Pernes, and Jaime S Cardoso. A robust fingerprint presentation attack detection method against unseen attacks through adversarial learning. In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE, 2020.

[6] João Afonso Pinto Pereira. Fingerprint anti spoofing-domain adaptation and adversarial learning. 2020.

[7] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL `http://arxiv.org/abs/1505.04597`.

[9] Ana F Sequeira, Shejin Thavalengal, James Ferryman, Peter Corcoran, and Jaime S Cardoso. A realistic evaluation of iris presentation attack detection. In *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, pages 660–664. IEEE, 2016.