Adversarial learning for a robust fingerprint presentation attack detection method against unseen attacks

João Afonso Pereira[1]
joao.p.pereira@inesctec.pt
Diogo Pernes[1,3]
dpc@inesctec.pt
Ana F. Sequeira[1]
ana.f.sequeira@inesctec.pt
Jaime S. Cardoso[1,2]
jaime.cardoso@inesctec.pt

[1] INESC TEC
Porto, Portugal
[2] Faculdade de Engenharia da Universidade do Porto
Porto, Portugal
[3] Faculdade de Ciências da Universidade do Porto
Porto, Portugal

Abstract

Fingerprint presentation attack detection (PAD) methods present a stunning performance in current literature. However, the *fingerprint PAD generalisation problem* is still an open challenge requiring the development of methods able to cope with sophisticated and unseen attacks as our eventual intruders become more capable. This work addresses this problem by applying a regularisation technique based on an adversarial training and representation learning specifically designed to improve the PAD generalisation capacity of the model to an unseen attack. The application of the adversarial training methodology is evaluated in two different scenarios: i) a handcrafted feature extraction method combined with a Multilayer Perceptron (MLP); and ii) an end-to-end solution using a Convolutional Neural Network (CNN). The experimental results demonstrated that the adopted regularisation strategies equipped the neural networks with increased PAD robustness. The CNN models' capacity for attacks detection in the unseen-attack scenario was particularly improved, showing remarkable improved APCER error rates when compared to state-of-the-art methods in similar conditions.

1 Introduction

Fingerprint presentation attack detection (FPAD) methods have been developed to overcome the vulnerability of fingerprint recognition systems (FRS) to spoofing. However, most of the traditional approaches have been quite optimistic about the behavior of the intruder, assuming the use of a previously known type of attack sample. This assumption has led to the overestimation of the performance of the methods, using both live and spoof samples to train the predictive models and evaluate each type of fake samples individually [10].

In this work, the *FPAD generalisation problem* is addressed by means of a regularisation technique designed to improve the generalisation capacity to unseen attacks in which the proposed model jointly learns the representation and the classifier from the data, while explicitly imposing invariance to the presentation attack instrument (PAI) types aka, 'PAI-species', in the high-level representations for a robust PAD method. The contributions of this work are then two-fold: 1) application of the adversarial training concept to the generalisation to unseen attacks problem in FPAD; and 2) evaluation of the adversarial training methodology in: i) combination of handcrafted features with a Multilayer Perceptron (MLP); ii) a Convolutional Neural Network (CNN) end-to-end solution. In this paper, this section summarises the proposed work and how it addresses the research question posed, section 2 presents the methodology, section 3 describes the experimental setup, section 4 presents the results and discussion and finally section 5 concludes the work.

2 Proposed Methodology

This work applies the methodology from Ferreira *et al*. [1] which was adopted in Pereira *et al*. [9] with the appropriate adjustments. The original method was presented by Ferreira *et al*. [2] in the context of sign language recognition, in an approach that builds on those initially introduced by Ganin *et al*. [4], for domain adaptation, and Feutry *et al*. [3], to learn anonymized representations. The underlying idea behind this approach is that, in order to generalise well to unseen attacks, the model should not specialize in discriminating any of the PAI species (PAISp) presented at training time and, therefore, the learned internal representations should be invariant to the PAISp. For this purpose, the model combines an adversarial approach with a species-transfer training objective The high-level architecture of the model is summarized in Fig. 1. It should be assumed that one has access to a labeled dataset $\mathbb{X} = \{\boldsymbol{X}_i, y_i, s_i\}_{i=1}^{N}$ of N samples, where $\boldsymbol{X}_i$ represents the i-th input sample, and y_i and s_i denote the corresponding class label (*bona fide* or *attack*) and the PAI species (only defined for attack samples), respectively. Let $\mathbb{X}^{bf}$ and $\mathbb{X}^{a}$ be these partitions of $\mathbb{X}$ for bona-fide and attack samples, respectively, and N^{bf} and N^{a} their respective cardinality.

The model comprises three main sub-networks: (i) an encoder network $h(\cdot;\theta_h)$ that receives input samples and maps them to a latent space; (ii) a *task-classifier* network $f(\cdot;\theta_f)$ which aims to distinguish attack and bona fide samples, mapping latent representations to the corresponding class probabilities; and (iii) a *species-classifier* network $g(\cdot;\theta_g)$ that receives latent representations from attack samples and aims to predict the corresponding PAI species. The species-classifier is trained to minimize the classification loss of the PAI-species. Simultaneously, the task-classifier and the encoder are jointly trained to minimize the classification loss between attacks and bona fide samples, while trying to keep the PAI-species classification close to random guessing. In addition to the adversarial training, a species-transfer objective is employed to further encourage the latent representations to be species-invariant. The overall objective function of the encoder and task classifier is then the combination of the previous objectives and can be formulated as:

$$\min_{\theta_h,\theta_f} \mathcal{L}(\theta_h,\theta_f,\theta_g) = \min_{\theta_h,\theta_f} \left\{ \mathcal{L}_{\text{task}}(\theta_h,\theta_f) + \lambda \mathcal{L}_{\text{adv}}(\theta_h,\theta_g) + \gamma \mathcal{L}_{\text{transfer}}(\theta_h) \right\}, \tag{1}$$

where $\gamma \geq 0$ is the weight that controls the relative importance of the species-transfer term and the objective for the species-classifier remains unchanged.

3 Experimental Setup

For more details, the reader is referred to Pereira *et al*. [9].

PAD Performance Evaluation Metrics: *Equal Error Rate (EER)*, *Attack Presentation Classification Error Rate* (APCER) and*Bona-fide Presentation Classification Error Rate* (BPCER) for APCER of 5% as in [6].

Dataset: The Fingerprint LivDet2015 [7] training dataset comprises a set of five subsets, each one corresponding to a specific fingerprint sensor. For each sensor there are bona fide samples and different types of PAI.

Evaluation protocols: The framework is denominated "unseen-attack", as the PAI seen in the testing phase is unknown to the model.

Handcrafted feature extraction method: Histograms of intensity, Local Binary Patterns (LBP and Local Phase Quantization .

Implementation details: The models were implemented in Python with the PyTorch library. For details, see Pereira *et al*. [9].

4 Results and discussion

In Table 2, the results of the baseline methods (*MLP* and *CNN*) and their respective regularised versions (MLP_{reg} and CNN_{reg}) are displayed. Comparing the performance of the baseline and regularised versions, it can be observed that: i) regarding the MLP, except for the Hi Scan sensor, in all the cases there is a significant improvement in at least 2 out of the 3 presented metrics; and ii) regarding the CNN, there is a significant improvement without exception in all error rates, with a particular significant improvement of the APCER value from 4.12% to 0.81% (for the average of the five sensors). From these observations, it can be stated with confidence that, overall, the regularisation technique improves the PAD robustness of both the models.

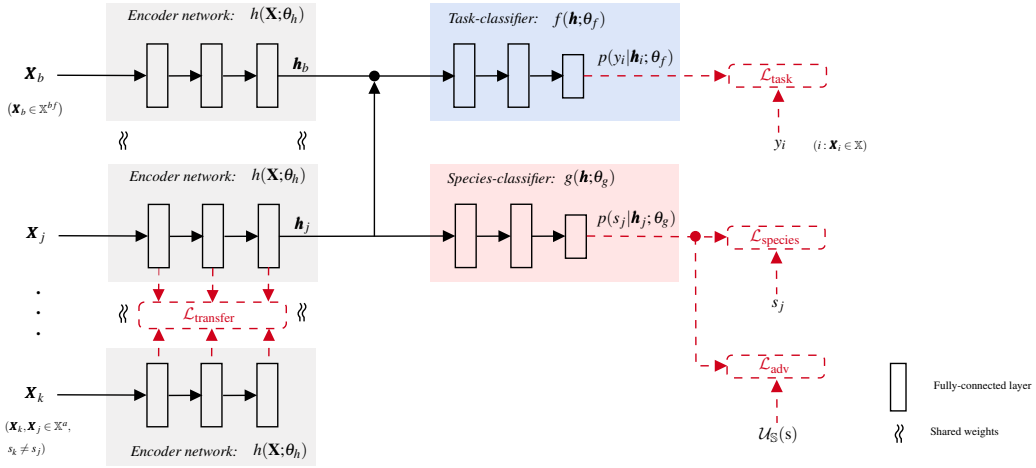

Figure 1: The architecture of the proposed species-invariant neural network (from [1]).

Still, it is arguable that the performance of the MLP, even the baseline version, outperforms the CNN results. Nevertheless, it should be noted that: i) the first scenario is taking advantage of rich handcrafted features; and ii) the data available for training is not enough to take the best out of the CNN learning capabilities. Thus, on the one hand the end-to-end solution provided by the CNN saves a considerable effort in the computation of the feature extraction step and, on the other hand, increasing the amount of training data will certainly increase the performance of these models, as there is a high potential for growth.

Table 1: Baseline and proposed regularised approaches - Cross Match, Digital Persona and Green Bit sensors. (*BPCER@APCER* = 5% noted by *BPCER@5*.)

Method	PAD metrics (%)								
	Cross Match			*Digital Persona*			*GreenBit*		
	APCER	BPCER@5	EER	APCER	BPCER@5	EER	APCER	BPCER@5	EER
MLP	**0.07**	7.57	4.33	0.00	0.53	0.45	**0.70**	**0.20**	1.10
MLPreg	0.13	**4.30**	**3.70**	**0.00**	**0.00**	**0.30**	**0.70**	0.63	**0.93**
CNN	5.00	6.25	8.70	5.60	10.80	7.28	3.03	14.13	7.05
CNNreg	**1.07**	**4.65**	**2.82**	**0.60**	**3.85**	**2.45**	**0.60**	**2.93**	**1.63**

Table 2: Baseline and proposed regularised approaches - Hi Scan and Time Series sensors, as well as the average of the results for the 5 sensors. (*BPCER@APCER* = 5% noted by *BPCER@5*.)

Method	PAD metrics (%)								
	Hi Scan			*Time Series*			Average of the 5 sensors		
	APCER	BPCER@5	EER	APCER	BPCER@5	EER	APCER	BPCER@5	EER
MLP	**0.30**	**2.83**	**3.03**	**0.00**	**0.03**	0.60	**0.21**	2.23	1.90
MLPreg	1.30	3.60	3.38	**0.00**	**0.03**	**0.10**	0.43	**1.71**	**1.68**
CNN	5.60	20.15	11.25	1.37	9.10	4.07	4.12	12.09	7.67
CNNreg	**1.20**	**1.21**	**1.04**	**0.60**	**6.30**	**2.70**	**0.81**	**3.79**	**2.13**

Despite the evidences showed in favour of the effectiveness of the regularisation technique, it is crucial to compare the results obtained with the proposed approach against the current state-of-the-art DL based PAD that tackle the unseen-attack scenario. This is not an easy task as most works still opt for a more traditional approach, based on binary classification limited to one type of attack at a time. From the available literature using similar databases and addressing the generalisation problem, stands out the meritory initiative of Fingerprint LivDet2015 [7] of evaluating the methods with some unseen types of PAISp.

Table 3 presents the results of the proposed regularised CNN version, *CNNreg*, alongside with the comparable literature methods currently available. The comparison shows the best results for common subsets of the used database presented in the LivDet2015 [5, 7] competition, as well as with an additional recent publication [8]. From the observed results, it is remarked the significant improvement of the *CNNreg* in two out of three sensors and undoubtedly when considering the average values. In particular, the *CNNreg* provided an APCER value of 0.76% against 2.09% and 6.33% of the other methods (for the average of the three sensors).

Table 3: Literature and proposed approach.(*BPCER@APCER* = 5% noted by *BPCER@5*.)

Method	PAD metrics (%)							
	Cross Match		*Digital Persona*		*GreenBit*		Average	
	APCER	BPCER@5	APCER	BPCER@5	APCER	BPCER@5	APCER	BPCER@5
Proposed CNNreg	1.07	4.65	**0.60**	**3.85**	**0.60**	**2.93**	**0.76**	**3.81**
LivDet2015 [5, 7]	1.68	≈ **0.80**	**0.60**	≈ 10.00	4.00	≈ 5.00	2.09	≈ 5.27
Park et al. [8]	**0.00**	-	11.00	-	8.00	-	6.33	-

5 Conclusions and future work

Comparing the baseline and regularised versions, it can be stated that, overall, the regularisation technique improves the PAD robustness of both the models. Despite the fact that the *MLPreg* fed with rich handcrafted features proved to be competitive, the fact is that *CNNreg* has more potential for growth and for increasing its performance in the future. The comparison of the proposed approach against the current DL based PAD methods that tackle the unseen-attack scenario, is not an easy task as most works still opt for a more traditional approach based on binary classification limited to one type of attack at a time. Still, from the comparison with the available literature using similar databases and addressing the generalisation problem, it is verified a significant superiority of the *CNNreg* in two out of three sensors and undoubtedly when considering the average values.

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020 and by Ph.D. Grant SFRH/BD/129600/2017.

References

[1] Pedro Ferreira, Ana F. Sequeira, Diogo Pernes, Ana Rebelo, and Jaime S. Cardoso. Adversarial learning for a robust iris presentation attack detection method against unseen attack presentations. In *Proceedings of the 18th BIOSIG*, 2019.

[2] Pedro M. Ferreira, Diogo Pernes, Ana Rebelo, and Jaime S. Cardoso. Learning signer invariant representations with adversarial training. In *12th ICMV*, 2019.

[3] Clément Feutry, Pablo Piantanida, Yoshua Bengio, and Pierre Duhamel. Learning anonymized representations with adversarial neural networks. *arXiv:1802.09386*, 2018.

[4] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proc. 32nd Int. Conf. ML*, 2015.

[5] Luca Ghiani, David A. Yambay, Valerio Mura, Gian Luca Marcialis, Fabio Roli, and Stephanie A. Schuckers. Review of the fingerprint livdet competition: 2009 to 2015. *IMAV*, 58:110 – 128, 2017.

[6] ISO/IEC JTC1 SC37. Information Technology - Biometrics - Presentation attack detection Part 3: Testing and Reporting. 2017.

[7] Valerio Mura, Luca Ghiani, Gian Marcialis, Fabio Roli, David Yambay, Schuckers, and Stephanie Schuckers. Fingerprint LivDet2015.

[8] E. Park, X. Cui, T. H. B. Nguyen, and H. Kim. Presentation attack detection using a tiny fully convolutional network. *IEEE TIFS*, 14 (11):3016–3025, 2019.

[9] Joao Pereira, Ana F. Sequeira, Diogo Pernes, and Jaime S. Cardoso. A robust fingerprint presentation attack detection method against unseen attacks through adversarial learning. In *19th BIOSIG*, 2020.

[10] Ana F. Sequeira and Jaime S. Cardoso. Fingerprint liveness det. in the presence of capable intruders. *Sensors*, 15:14615–14638, 2015.