Exposing Presentation Attacks by a Combination of Multi-intrinsic Image Properties, Convolutional Networks and Transfer Learning

Rodrigo Bresan, Carlos Beluzo, and Tiago Carvalho(✉)

Federal Institute of São Paulo, Campinas, SP 13069-901, Brazil
tiagojc@gmail.com

Abstract. Nowadays, adoption of face recognition for biometric authentication systems is widespread, mainly because this is one of the most accessible biometric characteristic. Techniques intended on deceive these kinds of systems by using a forged biometric sample, such as a printed paper or a recorded video of a genuine access, are known as presentation attacks. Presentation Attack Detection is a crucial step for preventing this kind of unauthorized accesses into restricted areas or devices. In this paper, we propose a new method that relies on a combination of the intrinsic properties of the image with deep neural networks to detect presentation attack attempts. Exploring depth, salience and illumination properties, along with a Convolutional Neural Network, proposed method produce robust and discriminant features which are then classified to detect presentation attacks attempts. In a very challenging cross-dataset scenario, proposed method outperform state-of-the-art methods in two of three evaluated datasets.

Keywords: Presentation attack · Spoofing attack · Transfer learning · CNN · Intrinsic image properties

1 Introduction

Biometrics consists in identify a given individual by its physiological traits (e.g., face, iris or fingerprint) or behavioral patterns (e.g., keystroke dynamics, gait) and it have been used on different types of devices for authentication purpose. Attacks to biometric systems are known as presentation or spoofing attacks. It consists in present a synthetic biometric sample, simulating biometric pattern of a valid user, to the system in order to obtain access as a legitimate user.

To fight back presentation attacks, different literature methods have been proposed in the last years. According to Pan *et al.* [10], techniques for Presentation Attack Detection (PAD) can be grouped into four major groups:

J. Blanc-Talon et al. (Eds.): ACIVS 2020, LNCS 12002, pp. 153–165, 2020.
https://doi.org/10.1007/978-3-030-40605-9_14

user behavior modeling, data-driven characterization, user cooperation and hardware-based.

Techniques based on behaviour modeling for PAD consists in models user's behaviors, such as head movements and eye blinking. Data-driven techniques are based on finding artifacts in attempted attacks by exploiting data that came from a standard acquisition sensor. User cooperation based techniques focus on interaction between user and authentication system, such as asking the user to execute some movements. Finally, there are techniques that use extra hardware, such as depth sensors and infrared cameras, to obtain more information about the scenario to finding cues that reveal an attempted attack[1].

Schwartz *et al.* [16] presented an anti-spoofing method by exploring the use of several visual descriptors for characterizing facial region according its color, texture, and shape properties. To deal with the high dimensionality in final representation vector, the authors proposed to use Partial Least Squares (PLS) classifier, an statistical approach for dimensionality reduction and classification, which was designed to distinguish a genuine biometric sample from a fraudulent one.

Pinto *et al.* [15] proposed a data-driven method for video PAD based on Fourier analysis in residual noise signature extracted from input videos. Use of well-known texture feature descriptors, such as Local Binary Patterns was also considered in the literature by Maata *et al.* [9], which focuses on detecting micro-texture patterns that are added into the fake biometric samples during the acquisition process. Approaches based on Differences of Gaussian (DoG) [12,18] and Histogram of Oriented Gradients (HOG) [7,19] were also proposed, but at the cost of final results is affected by illumination conditions and the capture sensor, due to their nature.

Yeh *et al.* [21] proposed an effective approach against face presentation attacks, based on perceptual image quality assessment, by adopting a Blind Image Quality Evaluatior (BIQE) along with a Effectivate Pixel Similary Deviation (EPSD), to generate new features to use on a multi-scale descriptor, showing it's efficacy when compared to previous works.

In this paper we introduce a new PAD technique which requires no additional hardware components (e.g., depth sensor, infrared sensor). Different intrinsic image properties are estimated and combined with a Convolutional Neural Network (CNN) and applying a transfer learning process we are able to extract robust and discriminative features. These features are then fed into a Extreme Gradient Boosting (XGBoost) classifier and a classification process with two steps is applied in order to classify samples into attack attempt or genuine sample.

Proposed method outperformed many existing literature approaches for face PAD problem, presenting better results in two of three datasets evaluated.

The main contributions of this paper include: (1) proposition of a new method for face PAD, which is based on a combination between

[1] Since this paper focus on data-driven techniques, we focused our literature review on this kind of methods.

intrinsic image properties and deep neural networks; (2) evaluation of different intrinsic properties (e.g., saliency, depth and illumination maps) for the PAD problem, which to the best of our knowledge, have never been evaluated in this context; (3) expressive results for both cross and intra dataset protocol in different public datasets; (4) effective application of transfer-learning approach in a PAD context.

2 Proposed Method

The method proposed in this paper can be divided in four main steps as depicted on Fig. 1. First state consists on estimate intrinsic properties from images. Then, we use a ResNet50 to extract bottleneck features which are submitted to the first classification step by an XGBoost classifier. This step calculates probabilities for each video frame to be, or not, part of an attack attempt. Then, these probabilities are used in a final stage, which performs a meta-learning process combining information from illumination, depth, and saliency maps, resulting in a new artifact, referred in this paper as fusion vector. Finally, this fusion vector feed a second XGBoost classifier responsible for the final prediction.

2.1 Intrinsic Images Properties Estimation

In order to extract intrinsic image information from video samples, for each frame, intrinsic image properties are extracted, which generates intermediate level image representations as depicted on Fig. 2.

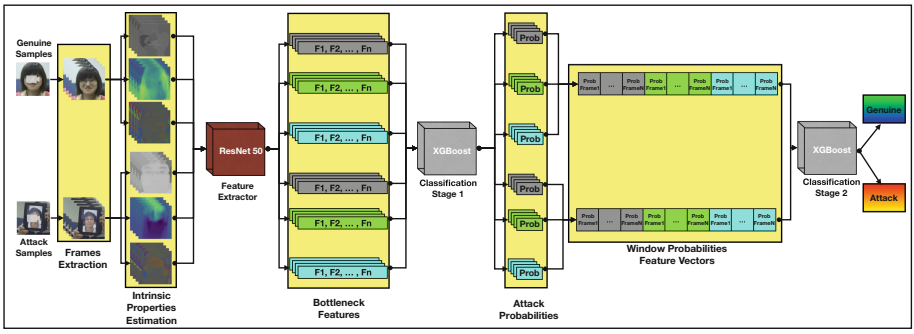

Fig. 1. Overview of the proposed method. Each video sample is split into frames and from each frame, intrisic image properties are calculated. Then, using a ResNet50, proposed method extracts bottleneck features, which are classified by an XGBoost according probability to be an attack. Probabilities of different intrinsic properties are then combined, by using a window of N frames, where N is the small number of frames in a video of evaluated datasets, into a final feature vector which is classified according it average probability of all frames.

Depth Maps. Due to the fact of presentation attacks being frequently reproduced over a flat surface, such as a sheet of paper or a tablet, we believe that the depth estimation from a given biometric sample can provide relevant information about its authenticity. Our hypothesis is that when presented with a flat surface, depth map estimated from a sample should differ from a real face.

Proposed method estimates depth maps by using Godard *et al.* [5] method, which uses stereo images to train a fully convolutional deep neural network associated with a modified loss function to estimates image depth. This trained network is then used to estimate depth maps from a single image. As described in Sect. 2.2, here we also take advantage of transfer learning approach, transferring weights from the method proposed by Godard *et al.* to our estimator.

Godard *et al.* method's learn a function f which can predict the depth from a given pixel on a single image. Using an unsupervised learning approach, the authors propose to reconstruct a given image from another, based on a calibrated pair of binocular cameras, thus allowing the learning of 3D cues of the original image. This is performed by finding depth field from the left image, and then reconstructing the correspondent right image. By using a modified loss function that outputs the disparity maps, which combines the smoothness, reconstruction and left-right consistency, the method estimates depth map from a single image.

Illumination Maps. In digital forensics, illumination inconsistencies have been frequently used to detect image forgeries [1,2]. Inspired by these works, proposed method also take advantage of illuminant maps to encode illumination information into PAD context. Our hypothesis is that generated illumination maps from a real face will show differences in its reflection when compared to the generated illumination map from a face depicted in a flat surface.

To capture illumination information, we calculate illuminant maps from each frame using the approach proposed by Riess and Angelopoulou [14]. This method estimates illuminant maps by using the Inverse Intensity-Chromaticity Space where the intensity $f_c(\mathbf{x})$ and the chromaticity $\chi_c(\mathbf{x})$ of a color channel $c \in \{R, G, B\}$ at position $\mathbf{x}$ is represented by

$$\chi_c(\mathbf{x}) = m(\mathbf{x}) \frac{1}{\sum_{i \in \{R,G,B\}} f_i(\mathbf{x})} + \gamma_c \ . \tag{1}$$

In Eq. 1, γ_c represents the chromaticity of the illuminant in channel c, whereas $m(\mathbf{x})$ mainly captures geometric influences, *i.e.* light position, surface orientation and camera position, and is approximate as described in [17].

Saliency Maps. As in depth and illumination cases, proposed method also takes advantage of saliency information using the same hypothesis that flat objects used in PAD will spoil quality in saliency estimation.

Saliency maps are estimated using the method proposed by Zhu *et al.* [24] which have two major steps: (1) a background modeling using boundary connectivity, which characterizes the spatial layout of image regions with respect to

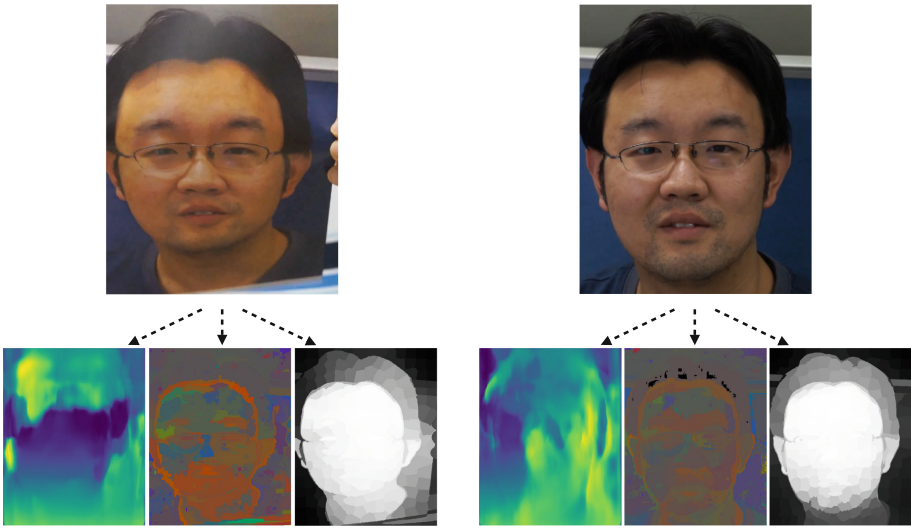

Fig. 2. Intrinsic image properties representation. Comparison between a presentation attack (upper left) with a genuine user (upper right). Below each picture is presented the generated map for depth, illumination, and saliency, respectively.

image boundaries; (2) a principled optimization framework to integrate multiple low-level cues, including proposed background measure. The following equation denotes the method proposed by Zhu *et al.* [24], to generate a saliency map from a single image.

$$BndCon(R) = \frac{|\{p|p \in R, p \in Bnd\}|}{\sqrt{|\{p|p \in R\}|}} \tag{2}$$

where p is a given image patch and Bnd is the set of image boundary patches.

2.2 Features Extraction

Once intrinsic image properties maps are estimated, next step of proposed method consists in extract features from each intermediate representation map. To accomplish this task, first we perform an alignment at eye's level on all of our frames and their property maps, followed by a crop on the face region, avoiding background and scene information[2].

Next, proposed method takes advantage of a combination between an well know CNN architecture and the transfer learning process [22]. We choose ResNet50 [6], a robust and effective CNN architecture, associated with ImageNet weights, to extract features from previously generated maps. Removing top layer, ResNet50 works as a feature extractor, which provides feature vectors commonly known as bottleneck features. As the final output of this step, a

[2] A classifier which consider scene information could lead to undesirable features and an unfair comparison against literature methods.

feature vector of 2,048 dimensions will be generated, which we will be later on referred to as the bottleneck feature vector.

2.3 Classification

Proposed method uses a two-stages classification pipeline, in which the first classifier is used for frames classification, while the latter one is used for classifying samples (videos) itself.

Stage 1. First stage use an XGBoost [3] classifier, due to its robustness in the task of binary classification when using multiple features. Given a bottleneck feature vector, our classifier returns for each frame, the probability of that frame belong to an attack video, or not. This stage results in 8 probabilities for each frame (probability to be an attack, or not, from frame itself, probability to be an attack, or not, from illuminant map, probability to be an attach, or not, from depth map and probability to be an attack, or not, from salience map).

Stage 2 (Fusion). Given an input video V_P, which already have intrinsic properties estimated, composed by n frames $f_1^P, f_2^P, \cdots, f_n^P$, and where P denotes the intrinsic property extracted from the video ($P \in \{\text{D, I, S}\}$). In previous stage, we estimated probability for each frame belonging to a class or another, denoted by f_i^P.

Using a fusion-based approach, we combine information from all intrinsic image properties in a way to use all these information together, resulting in a **P**robability **F**eature **V**ector (PFV) defined by

$$PFV = \{p^D, p^I, p^S\} \tag{3}$$

where p^P is given by

$$p^P = f_1^P, f_2^P, \cdots, f_m^P \qquad P \in \{D, I, S\} \tag{4}$$

where m is given by the number of frames into the video with small number of frames in dataset, D, I and S represents depth, illumination and salience maps, respectively.

Finally, PFV vectors are classified using a second XGBoost classifier.

3 Experiments and Results

To evaluate proposed method, different rounds of experiments were performed using three public anti-spoofing datasets, containing samples from genuine accesses and presentation attacks. The adoption of protocols focused in intra-dataset evaluation, where one dataset is tested within the same scenario was performed by following the protocols suggested by datasets' creators. Evaluation of different datasets scenarios, commonly known as inter-dataset or cross-dataset,

was also conducted, to assess the performance of proposed method in unknown scenarios. This latter one is the most challenging in the literature, due to the differences in capture conditions that one dataset shows from another one.

Furthermore, it is also paramount to realize that, since we are interested in evaluate the efficiency of each intrinsic property individually, final results reported for depth, illumination, and saliency reflects a majority vote process among all the frames classified on Stage 1.

3.1 Datasets, Metrics, and Setup

To address the efficiency of the proposed method, three publicly available anti-spoofing datasets were selected. The criteria for selection of these datasets among many others available was due to their major adoption in previous works that tackle PAD.

Replay-Attack [4]**.** Consisting of 1300 video clips from both photo and video attacks from 50 subjects, the Replay-Attack (RA) dataset shows itself as a reliable dataset for the evaluation of the hereby proposed method, once it is presented with different lighting and environmental conditions. In this dataset, three different types of attack are provided: print attacks, mobile attacks, and video attacks. It is separated into three subsets: training set (containing 360 videos); development set (containing 360 videos); testing set (containing 480 videos); and enrollment set (containing 100 videos);

CASIA-FASD [23]**.** The CASIA-FASD dataset contains a total amount of 600 videos from 50 different subjects, created to provide samples from many of the existent types of presentation attacks. The videos are presented in twelve different scenarios, where each of them is composed by three genuine accesses and three attacks from the same person. Three different resolutions were used to capture (low, normal and high), along with three different types of attack (normal, printed attacks, printed and warped, printed with cut on the eyes region and video-based attacks).

NUAA Photograph Imposter Dataset [18]**.** The NUAA Photograph Imposter Dataset is composed of 15 subjects, comprising a total of 5,105 valid access images and 7,509 presentation attacks collected through a generic webcam at 20 fps with a resolution of 640 $\times$ 480 pixels. The subjects were captured over three sections in different places and lighting conditions. The production of the attack samples was made by shooting a high-resolution photograph with a Canon digital camera.

Metrics. To allow the comparison of the results obtained in this work, we adopt the *Half Total Error Rate* (HTER), which is measured by the mean value between the False Acceptance Rate (FAR), denoted by the rate of attack attempts misclassified as authentic, and the False Rejection Rate (FRR), which is denoted by the rate of authentic samples misclassified as attack. The HTER is measured by

$$\text{HTER} = \frac{\text{FAR} + \text{FRR}}{2} \tag{5}$$

where FAR is the False Acceptance Rate and FRR is the False Rejection Rate.

Experimental Setup. For illumination maps and its segmentation, parameters are the same as the presented in the work of Carvalho *et al.* [2]. For the depth and saliency maps, proposed method uses default parameters as suggested by Godard *et al.* [5] and Zhu *et al.* [24], respectively.

For Stage 1 and Stage 2, classification steps, proposed method uses XGBoost with a *gamma* of 0, a *max_depth* of 6, *gbtree* as booster and a learning rate of 0.3.

Experiments have been conducted by using Python programming language (version 3.6), along with the Keras[3] (version 2.2) and TensorFlow[4] (version 1.8).

3.2 Intra-dataset Evaluation

In intra-dataset evaluation evaluation protocol, we apply the same protocols proposed by each databases' authors, and use HTER metric to measure performance.

As displayed in Table 1, the usage of the fusion outperformed single properties results in Replay Attack, with an HTER value of 3.75%. For CASIA dataset, best results have been achieved using fusion, yielding an HTER result of 9.63%. Finally, results in NUAA dataset using depth maps outperformed all the other features, yielding an HTER of 18.31%.

Table 1. Results (in %) considering the Intra-Dataset protocol for the RA, CASIA and NUAA datasets.

Method	RA HTER	CASIA HTER	NUAA HTER
Raw	6.00	15.74	26.35
Depth	30.25	44.44	**18.31**
Illumination	16.12	16.11	43.65
Saliency	18.37	29.25	31.24
Fusion	**3.75**	**9.63**	26.34

These results present the importance of individual features and increase our hypothesis that different intrinsic properties can be used together to detect attacks. In special, depth maps depicted special representation value in attack detection process.

[3] https://keras.io.
[4] https://www.tensorflow.org.

3.3 Cross-Dataset Evaluation

Building a method that is highly adaptable from one face anti-spoofing database to another unknown one has been posed as a major challenge in previous works, and it's an essential ability for real-world applications that rely on face recognition for authentication.

This experiment presents results for the cross-dataset (inter-dataset) evaluation protocol, when one dataset have been used for training while a different one have been used for testing. Table 2 present results when testing method over RA, CASIA and NUAA datasets, respectively.

Table 2. Results (in %) considering the Cross-Dataset Protocol using as test dataset RA (left), CASIA (middle), and NUAA (right).

Train/Test Set	Method	HTER
NUAA/RA	Raw	57.14
	Depth	**49.00**
	Illumination	56.28
	Saliency	62.92
	Fusion	58.64
CASIA/RA	Raw	51.57
	Depth	55.71
	Illumination	**45.21**
	Saliency	48.42
	Fusion	46.71

Train/Test Set	Method	HTER
NUAA/CASIA	Raw	38.33
	Depth	44.81
	Illumination	54.07
	Saliency	48.33
	Fusion	**35.37**
RA/CASIA	Raw	55.55
	Depth	51.11
	Illumination	50.92
	Saliency	**50.74**
	Fusion	59.44

Train/Test Set	Method	HTER
CASIA/NUAA	Raw	38.13
	Depth	**34.11**
	Illumination	50.22
	Saliency	48.37
	Fusion	35.67
RA/NUAA	**Raw**	**51.67**
	Depth	60.35
	Illumination	52.21
	Saliency	58.18
	Fusion	51.88

From presented tables is not difficult to realize that different intrinsic help in different ways for cross-dataset scenario. This fact expose that different kinds of intrinsic properties collaborate differently for each scenario but always aggregating some important information.

Again, better HTERs are achieved when using Depth (training on CASIA dataset and testing on NUAA dataset) and Fusion approaches (training on NUAA dataset and testing on CASIA).

3.4 Comparison with State-of-the-art

Since cross-dataset represents more challenging scenario, this experiment compares achieved results against some state of the art methods. Table 3 summarize best results (HTER) obtained for proposed method compared against some state-of-the-art methods.

When compared against state-of-the-art methods, proposed method outperformed literature in two of three datasets for cross-dataset protocol. Testing on NUAA dataset, proposed method achieved an HTER value of 34.11% when trained on the CASIA dataset, outperforming results obtained in previous works [12,18]. For the CASIA dataset, the best results were attained with the usage of the features fusion, with an HTER of 35.37% when trained on NUAA dataset. The best results for the RA dataset were achieved by the usage of the illumination maps, with an HTER of 45.21%, but outperformed by Yang *et al.* [20].

Table 3. Comparison among existing approaches for cross-dataset evaluation protocol.

Method	CASIA	RA	NUAA
Yeh *et al.* [21]	39.00	38.10	–
Pinto *et al.* [13]	47.16	49.72	–
Yang *et al.* [20]	42.04	41.36	–
Patel *et al.* [11]	–	**31.60**	–
Tan *et al.* [18]	–	–	45.85
Peixoto *et al.* [12]	–	–	49.85
Raw image	38.33	51.57	38.13
Depth	44.81	49.00	**34.11**
Illumination	50.92	45.21	50.22
Saliency	48.33	48.42	48.37
Fusion	**35.37**	46.71	35.67

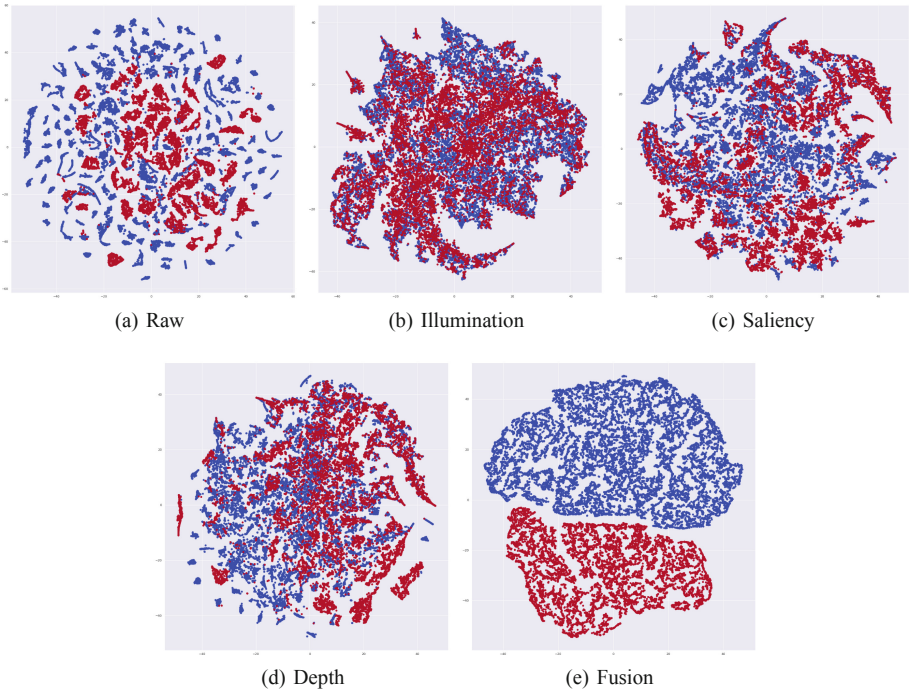

Fig. 3. t-SNE features extracted from Replay Attack dataset. Each figure depicts features for an specific intrinsic properties, where blue points represents genuine access samples and red points represent attack samples. Each intrinsic property perform a different degree of separability between samples. Fusion of all of the intrinsic features perform a considerably separability between classes. (Color figure online)

3.5 Intrinsic Properties and Features Analysis

Last experiment performed on proposed method focus on show how each one of intrinsic properties contribute to improve classes separability. This analysis is performed using T-distributed Stochastic Neighbor Embedding (tSNE) [8], which project into a 2D feature space bottleneck features (originally with 2048 dimensions) extracted from each intrinsic property map. Figure 3 depicts feature vectors extracted from Replay Attack dataset. Each figure depicts features for an specific intrinsic properties, where blue points represents genuine access samples and red points represent attack samples. Each intrinsic property perform a different degree of separability between samples. Fusion of all of the intrinsic features perform a considerably separability between classes.

4 Conclusions and Research Directions

In this paper, we have proposed a new method that, by using a two-step classification model, along with intrinsic image properties, such as depth, illumination, and saliency, learn representative features for the task of presentation attack detection. Evaluating the hereby proposed method in three different databases, we reach results outperforming previous works for PAD problem. Findings provided by this paper, such as the efficacy of using image intrinsic properties, can lead to a better understanding on the study and development of new anti-spoofing methods, as well as to provide better insights for development of new datasets. Our results also confirm our hypothesis that by adopting transfer learning techniques along intrinsic image properties, are capable to detect attempts of presentation attacks.

For future works, we intend to investigate other types of intrinsic properties, to better understand the features that may help in the task of distinguishing between an authentic facial biometric sample and a fraudulent one. We also believe that by performing a finetuning step, we could achieve even better results, once that the results attained in this work were achieved by adopting the weights of a pretrained network on data that does not share many similarities with the problem of PAD.

Acknowledgments. We would like to thank São Paulo Research Foundation (FAPESP) (#2017/12631-6), to the National Council for Scientific and Technological Development - CNPq (#423797/2016-6), and to NVIDIA for the donation of a TITAN XP GPU to be used on this research.

References

1. Carvalho, T., Faria, F.A., Pedrini, H., da Silva Torres, R., Rocha, A.: Illuminant-based transformed spaces for image forensics. IEEE Trans. Inf. Forensics Secur. **11**(4), 720–733 (2016). https://doi.org/10.1109/TIFS.2015.2506548

2. de Carvalho, T.J., Riess, C., Angelopoulou, E., Pedrini, H., de Rezende Rocha, A.: Exposing digital image forgeries by illumination color classification. IEEE Trans. Inf. Forensics Secur. **8**(7), 1182–1194 (2013). https://doi.org/10.1109/TIFS.2013.2265677
3. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ACM (2016)
4. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. In: 2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), pp. 1–7, September 2012
5. Godard, C., Aodha, O.M., Brostow, G.J.: Unsupervised monocular depth estimation with left-right consistency. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6602–6611, July 2017. https://doi.org/10.1109/CVPR.2017.699
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Komulainen, J., Hadid, A., Pietikinen, M.: Context based face anti-spoofing. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–8, September 2013. https://doi.org/10.1109/BTAS.2013.6712690
8. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
9. Maatta, J., Hadid, A., Pietikinen, M.: Face spoofing detection from single images using micro-texture analysis. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–7, October 2011. https://doi.org/10.1109/IJCB.2011.6117510
10. Pan, G., Wu, Z., Sun, L.: Liveness detection for face recognition. In: Delac, K., Grgic, M., Bartlett, M.S. (eds.) Recent Advances in Face Recognition, chap. 9, pp. 235–252. IntechOpen, Rijeka (2008). https://doi.org/10.5772/6397
11. Patel, K., Han, H., Jain, A.K.: Cross-database face antispoofing with robust feature representation. In: You, Z., et al. (eds.) CCBR 2016. LNCS, vol. 9967, pp. 611–619. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46654-5_67
12. Peixoto, B., Michelassi, C., Rocha, A.: Face liveness detection under bad illumination conditions. In: 2011 18th IEEE International Conference on Image Processing, pp. 3557–3560, September 2011. https://doi.org/10.1109/ICIP.2011.6116484
13. Pinto, A., et al.: Counteracting presentation attacks in face, fingerprint, and iris recognition. In: Deep Learning in Biometrics, p. 245 (2018)
14. Riess, C., Angelopoulou, E.: Scene illumination as an indicator of image manipulation. In: Böhme, R., Fong, P.W.L., Safavi-Naini, R. (eds.) IH 2010. LNCS, vol. 6387, pp. 66–80. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16435-4_6
15. da Silva Pinto, A., Pedrini, H., Schwartz, W., Rocha, A.: Video-based face spoofing detection through visual rhythm analysis. In: 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images, pp. 221–228, August 2012. https://doi.org/10.1109/SIBGRAPI.2012.38
16. Schwartz, W.R., Rocha, A., Pedrini, H.: Face spoofing detection through partial least squares and low-level descriptors. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–8, October 2011. https://doi.org/10.1109/IJCB.2011.6117592

17. Tan, R.T., Ikeuchi, K., Nishino, K.: Color constancy through inverse-intensity chromaticity space. In: Digitally Archiving Cultural Objects, pp. 323–351. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-75807_16
18. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 504–517. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15567-3_37
19. Yang, J., Lei, Z., Liao, S., Li, S.Z.: Face liveness detection with component dependent descriptor. In: 2013 International Conference on Biometrics (ICB), pp. 1–6, June 2013. https://doi.org/10.1109/ICB.2013.6612955
20. Yang, J., Lei, Z., Li, S.Z.: Learn convolutional neural network for face anti-spoofing. arXiv preprint arXiv:1408.5601 (2014)
21. Yeh, C.H., Chang, H.H.: Face liveness detection based on perceptual image quality assessment features with multi-scale analysis. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 49–56. IEEE (2018)
22. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems, pp. 3320–3328 (2014)
23. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: 2012 5th IAPR International Conference on Biometrics (ICB), pp. 26–31, March 2012. https://doi.org/10.1109/ICB.2012.6199754
24. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2814–2821, June 2014. https://doi.org/10.1109/CVPR.2014.360