Transfer Learning for Face Anti-Spoofing Detection

Sandoval Veríssimo, Guilherme Gadelha, Leonardo Batista, *Member, IEEE*, João Janduy and Fábio Falcão

***Abstract*—In recent years, the demand for facial biometric authentication services has increased dramatically. Also, the efforts to cheat this type of system have become more common. In this paper, we propose a single shot CNN-based solution for the face anti-spoofing problem. We trained a deep learning model using transfer learning from a pre-trained VGG16 model. After some pre-processing we rely solely on the network to classify an image. We evaluate several implications of the preprocessing of data, investigate the implications of different amounts of background included in the picture, and the effect of data subsampling. Additionally, we analyze what happens when we sub-sample the training data. We evaluate our results in four publicly available datasets, drawing some insights on the results by using the Grad-CAM algorithm. Our approach is competitive when compared with similar methods. Moreover, we achieved our results while training with a fraction of the original datasets, enforcing that experiments can be run much quicker without sacrificing accuracy.**

***Index Terms*—Face Anti-Spoofing, Transfer Learning, Deep Learning, VGG16**

I. Introduction

Due to worldwide health reasons [1], [2], social distancing and isolation started to be encouraged [3]. Biometric authentication is then in high demand [4]. Face anti-spoofing is essential [5] to ensure secure facial recognition, facial detection, and other biometric verification systems. A "presentation attack" [6] is the presentation of an artifact or human characteristic to the biometric capture subsystem in a way that could interfere with the intended security policy of the biometric system. Often, attackers use images from social networks as a form of presentation attack.

Due to the recent rise in the popularity of biometric authentication systems, presentation attacks (or spoofing) have also grown in number [7]. Therefore, it is essential to combine both the biometric authentication system with some form of anti-spoofing detection. Presentation attacks are often performed in 3 main approaches: *Photo attacks*, where the attacker uses a photo of the person to fool the system; *Video replay attack*, where a pre-recorded video is used to simulate a live person, and *3D mask attacks*, where the attacker relies on 3D masks to try to fool the system. The first two methods are more commonly used since it is costly to get a 3D mask.

To combat these attacks, we can use either passive or active methods. There are several ways to perform passive presentation attack detection (PAD), such as motion, texture, or reflectance analysis. Each of these approaches has advantages and disadvantages. On active spoofing detection, the system requests input action from the user, such as looking left or looking right. This active spoofing detection works better against pre-recorded videos or photo attacks; however, 3D masks can bypass it to a certain degree.

Some approaches depend on specific hardware to gather more information to identify an attack. For example, thermal cameras can easily spot temperature differences between a person and a device. Additionally, we noticed that hardware-dependent methods are often combined [8], [9] with image-based methods. That said, that hardware is usually not available for the general public, making it less accessible and unfit for general purpose use.

In this work, we focus on face anti-spoofing using only a monocular camera. In image processing, Convolutional Neural Networks (CNNs) have gained increased popularity in the last decade. CNNs can learn features to identify an attack during its training process [10], [11]. In this paper, we propose implementing a single shot CNN-based solution for the face anti-spoofing problem. The single-shot approach means that only one photo is required to identify an attack instead of a sequence of video frames. Inspired by the work of [12] we also employ transfer learning on a VGG model. However, differently from their work, we do not include a PCA and an SVM to aid in the classification. Instead, we rely on the network alone to make all the decisions leading up to the classification.

The methodology used in this paper focuses on a data-centric approach. We evaluate several implications of the preprocessing of data, investigate the best amount of background to be included in the picture, the effect of data augmentation, and investigate the differences between different capture devices. Additionally, we analyze what happens when we sub-sample the training data. To the best of our knowledge no other paper has commented on the subsampling matter.

We evaluate our results in four publicly available datasets: OULU-NPU [13], Replay Attack [14], MSU-FASD [15] and NUAA [16], comparing the results with the ones presented in the literature. We then draw some insights using the Grad-CAM [17] algorithm. Our best results have achieved less than 0.2% EER in a specific dataset.

II. Related Works

This section will review some of the most common face anti-spoofing methods, focusing on general consumer devices. It is worth noting that these methods can be combined in a single system in a real case scenario. Such combinations are made so that some methods strengths can compensate for other methods weaknesses.

Sandoval Veríssimo, Vsoft, e-mail:sandoval.sousa@vsoft.com.br
Guilherme Gadelha, Vsoft, e-mail:guilherme.gadelha@vsoft.com.br
Leonardo Batista, UFPB, e-mail:leonardo@ci.ufpb.br
João Janduy, Vsoft, e-mail:jjanduy@vsoft.com.br
Fábio Falcão, Vsoft, e-mail:fabio.franca@vsoft.com.br

A. Liveness Cue-Based Methods

Liveness cue-based methods attempt to detect variations on the image that can indicate liveness, such as movement, eyes blinks, change in facial expression, and even micro-intensity variations of the blood pulse. These methods can be used both in an intrusive or non-intrusive way. In the first approach, a specific action is required from the user, whereas the last approach can be performed without user knowledge. Earlier works on non-intrusive methods used frequency-based features [18], 3D maps to estimate head motion [19], optical flow lines [20], eye blinks [21] and others. Intrusive works usually rely on a command-and-response system, in which the user has to match a certain action, such as saying a random sequence of numbers [22].

B. Texture Based Methods

Texture-based methods are the most used for the Face Anti-Spoofing problem. These are inherently non-intrusive methods that perform well on photo and display attacks.

Static texture-based methods rely on a single image to detect an attack. These include methods such as light reflectivity difference [18], light reflectivity with Difference of Gaussian (DoG) filtering [16], Contrast Limited Adaptive Histogram Equalization (CLAHE) [23], Local Binary Patterns (LBP) [24], Gabor Wavelets [25] and Histogram of Oriented Gradients (HOG) [26].

On the other hand, **dynamic methods** use temporal information of a sequence of images. Works also include the use of LBP [27], Histogram of Oriented Optical Flows (HOOF) [28], Fourier residual noise video analysis [29]. More recent works make use of CNNs; we detail these in subsection II-D.

C. 3D Geometric Cue-Based Methods

3D geometric cue-based methods use 3D geometric features to differentiate between a spoof and a genuine user. Approaches focus on the 3D reconstruction of the face based on a 2D photo [30] or the depth estimation of the image. The latter can more easily be done with specific 3D cameras; however, since this is not available for the general public, other methods are needed to predict the depth of a single image. Works such as [31] make use of a pseudo depth map of an image on a sequence of images.

D. CNN-Based Methods

Although they can also be used in texture-based methods and 3D pseudo-depth-based methods, we decided to dedicate a specific subsection to CNN-based approaches because our work focuses on a CNN solution.

The first work to employ CNNs on the face anti-spoofing problem [10], used a simple AlexNet with an SVM classifier at the end. The authors also confirmed a previous hypothesis that enlarging the face bounding box to include more background was beneficial for the model.

In [32], a proposed solution with the same network but without the SVM classifier was made. The work also proposed a voting system between a model trained on aligned faces and faces with the background.

The work in [12] proposed to make use of transfer learning into a model of the VGGFace network with images of the face spoofing problem. While methodologically similar to one of our approaches, the model itself would not classify the images. Instead, the work used CNN to extract a single fused feature and then use it on an SVM to make the classification.

In [33], a model is trained using two loss functions calculated on the final feature map, before the and classification result of the network was outputted. The losses were calculated on a per-pixel level, and results were promising.

More recent works such as [34] and [35] focus their approaches in domain generalization and domain invariant feature alignment. However none of these works use the same methods and the same datasets as our work, thus a fair comparison is not possible

III. Methodology

In this section, we detail the datasets used for training and evaluation. Also, we explain the preprocessing steps of the input images that we experimented and the impact they had on training and results. And finally we discuss the selected network architecture and the metric used for evaluation.

A. Datasets

1) Replay Attack: The Replay Attack dataset [14] is well known and vastly used in Face Anti Spoofing works. It consists of 1200 videos, where 200 are from real people, and 1000 are from attacks. These videos are divided among 50 identities with 15,15,20 identities for train, validation, and test, respectively. There is no overlap in identities between the different partitions—genuine and attack videos average around 15 and 9.5 seconds of duration, respectively. After extracting all the frames, we ended up with around 75000 authentic images and 235000 attack images. The videos were recorded in two ways: a controlled scenario with a uniform background, and a less-controlled scenario where the person stands in front of a less uniform background. The attacks were made using prints, photos on display, and video recordings.

2) NUAA: The first public dataset for the face anti-spoofing problem. The NUAA [16] dataset has around 12000 images of 16 individuals. By default, the dataset has only a train and test split. There is also an overlap in identities between the sets. The attacks on the dataset are only made with printed photos. The capture sessions were made in three different environments, thus changing the background between sessions. Not all identities have real and attack photos in all three sessions. There is also no pattern on how many images of each subject or each session exists.

3) MSU-MFSD: With 280 videos, 70 of them of genuine users and 210 of attacks, the MSU-MFSD [15] dataset is another publicly available dataset. There are a total of 35 identities in this dataset. These identities are divided initially into 15 for training and 20 for testing. The frame rate and length of the videos vary. After extracting the frames, we ended up with approximately 19000 authentic images and

58000 attack images. The attacks consist of printed photos and videos. This dataset was the first to introduce a mobile scenario on the attacks (previous attacks were replayed usually on iPad devices).

4) OULU-NPU: One of the newer datasets, OULU-NPU [13] was released in 2017. The dataset has 55 identities organized in 20 for training, 15 for validation, and 20 for testing. In total, there are 4950 videos, of which 990 are from real people. After extracting all the frames, we had around 130000 authentic and 530000 attack images. The amount of images in this dataset alone is greater than all the other datasets combined. Thus, another use of subsampling is to assure all the datasets are equally represented during the training process. The dataset mainly focuses on mobile devices cameras as the capture devices. The attacks are made with both print and replay methods.

B. Data Preprocessing

Previous works on the face anti-spoofing problem often use data preprocessing techniques to leverage models results. Our work focused mainly on four aspects: face cropping, face alignment, dataset subsampling, and data augmentation. Details and explanations on why we decided to do these preprocessing operations are given in the subsections below.

1) Face Cropping: In the face anti-spoofing domain, preprocessing the images is commonly done to standardize the data the model receives [11]. The network can focus on learning patterns in the same vicinity with fewer variations between faces. The simplest way to do this is by cropping all the region that is not the face. The first preprocessing step we did was to use the Dlib [36] face detector to detect the faces in the images, and then we cropped only the faces and discarded the rest of the image.

There are a few problems with this approach, however. First, cropping the image does not mean that it will keep the same width x height ratio, as the face bounding boxes may or may not be square. Since the input of the VGG model is 224x224, we often distort the image and stretch it in some direction. Further explanation of the results will be shown in section IV.

2) Face Alignment: Face alignment starts by identifying the structure of the human face with any face detector algorithm or technique. Then, we can attempt to align the face using translation, scale, and rotation operations. Doing so allows all the faces to be standardized based on a common point, such as eyes localization, as done by [37]. It also solves the simple crop's stretching problem since the geometric operations are done on expected output.

On the other hand, with both the crop and face alignment, background is lost. Background information can have features that identify an attack, such as fingers holding a photo or the borders of a display. Nevertheless, allowing background in the image allows for more varied scenarios to be seen during the training process; this can hamper the model's accuracy.

Considering this trade-off, we decided to experiment with different alignments. Using the Dlib face detector, we have information about the eyes locations. We then rotate the image according to the line between the eyes; then we align the faces on the central point between the eyes. After that, we set different fixed distances between the eyes (in pixels). This work evaluates the effects of alignments with five different distances between the eyes: 125, 100, 75, 67, and 50 pixels. We choose this values based on preliminary experiments, and found that these values were enough to understand the effect of different distances between the eyes on each dataset.

3) Subsampling: Due to the high frame rate of capture devices, there is not much variation between two frames. In order to reduce the amount of redundant information on the training and validation sets, we examined the implications of subsampling the datasets. Thus, we can compare the results and evaluate if our model can learn with fewer training images but more significant variations.

The subsampling procedure was done with the Structural Similarity Index Measure (SSIM) [38]. This algorithm is used to measure the similarities between two images. The number of images we kept at the end of the subsampling process was 12000 for each dataset. We chose this number since the smallest dataset (NUAA) has approximately this amount of images. Thus, all the datasets should have the same importance on the training process. We were also careful to keep the number of images per subject roughly the same in order to not incentivize any bias on a few identities. It is worth noting that since the NUAA dataset is already the smallest one in our experiments, this dataset is not being subsampled.

C. Network And Transfer Learning

Training CNNs from scratch can be a very time-consuming process. Depending on the amount of training data, converging into a robust CNN can take days even if top-quality hardware is available. One solution to reduce the training time is to use Transfer Learning from a network previously trained on a similar domain [39]. For example, we may use the knowledge acquired from a network trained for images classification - and later applied to facial recognition problems - like VGG [40] and apply it to the Face Anti-Spoofing domain with some adaptations of the network architecture.

Inspired by the work of Li et al. [12], we decided to use the VGG architecture as our base model, in our case, the VGG16. The VGG16 architecture [41] was proposed to classify the ImageNet dataset [42]. The dataset consists of over 14 million images divided into 1000 classes. The original VGG16 work was able to achieve over 92% top-5 accuracy. The model was trained for weeks using NVIDIA Titan Black GPU. The architecture comprises six blocks, five of them consisting of a couple of convolutional layers followed by a pooling layer. The final block consists of three different dense layers leading up to a softmax classification layer. Figure 1 [43] shows a scheme of the VGG16 model. It is worth noting that the decision to use the VGG16 model was taken after some preliminary experiments which showed that the transfer learning to a VGG16 model achieved better results than to the InceptionV3, MobileNetV2 and VGG19 architectures.

D. Training Parameters

We also defined a few hyperparameters for the training process, along with performing a transfer learning to the face

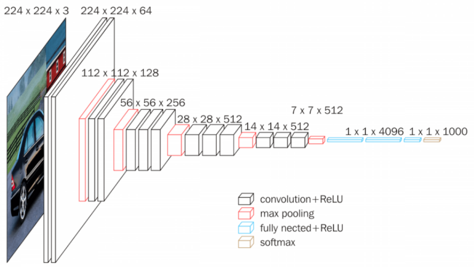


Fig. 1. VGG16 architecture scheme.

anti-spoofing domain on a VGG16 network. We added one dense layer before the classification layer. These two layers were the only ones to be trained from scratch, the base model weights were not modified. Table I shows the hyperparameters used in the training process.

TABLA I
TRAINING PARAMETERS USED ON THE NETWORK

Hyperparameter	Value
Optimizer	Adamax
Batch Size	16
Learning Rate	0.001
Dense Units	128
Dropout	0.3
# Epochs	10

These values were chosen after some initial experiments in a sample of the total amount of images. We varied each of these hyperparameters and studied the effects this change would imply on the accuracy of the model.

E. Metrics

We use the Equal Error Rate (EER) metric, which is a standard metric in the field [11]. The EER is defined as the point where False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal, considering a range of thresholds - usually between 0.0 and 1.0. This threshold determines if an image is an attack or not based on the comparison with the neural network output value for that specific image. Figure 2 [44] shows the identification of the EER point.

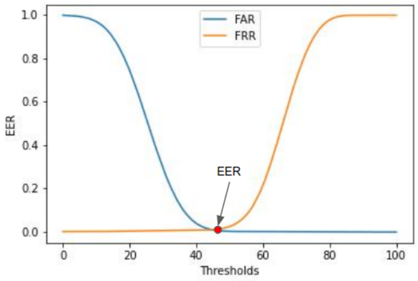


Fig. 2. Equal Error Rate identification

IV. EXPERIMENTS AND RESULTS

Below we discuss the experiments and results. For each experiment, we trained a network with the parameters defined in section III-D.

A. Face Cropping

We started our experiments evaluating the impact of face cropping in each dataset individually. After evaluating the NUAA, MSU, Replay Attack and OULU datasets, we noticed a stretching image issue. As explained before, this happens because cropped images need to be resized to the 224x224 input size of the VGG16 network, and as such, distortions would often happen. Each experiment model was trained and evaluated on a single dataset.

TABLA II
EER ON RAW VS CROPPED DATASETS.

Test Dataset	Raw Images	Cropped
NUAA	0%	4.18%
Replay Attack	0.67%	3.15%
MSU-FASD	5.32%	8.83%
OULU	4.69%	14.98%

Table II shows a comparison between the results of our method on both cropped and raw images. We can observe that the model achieved a worse performance across all datasets individually evaluated. These results suggest that non-facial features on the image, such as background, may contain valuable information to differentiate as spoofing attack from a real user.

Even though all models performed worse on the cropped version of the dataset, the impact of the crop was different in each one. For the NUAA dataset in particular the EER which was 0% reached over 4%. Looking at the images of the dataset, we believe that since most of the NUAA images are from people holding printed photos, the information of the borders of the paper and the fingers of the impostors was important in identifying some attacks.

B. Subsampling

Along with experimenting with the crop, we also analyzed the impact of the subsampling. The main purpose of these experiments was to understand how much the subsampling of training and validation sets affects the overall performance. Similarly to the face cropping experiments, we trained with all datasets (NUAA, MSU, Replay Attack and OULU) individually and tested in each one separately. The raw images experiments have no preprocessing as facial cropping, our goal is analyze these two preprocessing steps in parallel. The results are shown in Table III.

In Table III, the EER was higher for the Replay Attack and OULU datasets. Moreover, the results show no change in the EER for the NUAA dataset, and a 0.04% difference for the MSU-FASD dataset. We believe that this happened because all the datasets now have roughly the same proportions during the training process; therefore, the network learned more features

TABLA III
EER ON RAW VS SUBSAMPLED DATASETS.

Test Dataset	Raw Images	Subsampled
NUAA	0%	0%
Replay Attack	0.67%	1.98%
MSU-FASD	5.32%	5.36%
OULU	4.69%	7.18%

from the MSU-FASD and NUAA datasets and was less biased towards the OULU and ReplayAttack datasets.

It is also noteworthy that the two smallest datasets did not have any or had very little change in the EER. For the NUAA dataset kept the same 0% EER. This is expected, since this dataset was not subsampled at all. For the MSU-FASD dataset, a difference of 0.04% happened, a 0.75% increase on the overall EER of the first experiment.

C. Face Alignment

Again, in this case, each experiment used a single dataset for training and testing, i.e. following an intra-dataset evaluation protocol [11]. As we explained in Section III-B2, we experimented with five different eye distances. The results can be seen in Table IV.

TABLA IV
EER PER DATASET BY VARYING DISTANCE BETWEEN THE EYES (IN PIXELS).

Dataset	125px	100px	75px	67px	50px
NUAA	13.32%	5.30%	2.40%	1.75%	0.10%
Rep. Attack	2.72%	4.45%	5.53%	4.44%	12.86%
MSU-FASD	12.53%	15.93%	12.00%	12.82%	11.77%
OULU-NPU	14.97%	14.62%	13.44%	12.65%	14.01%

These results show that each dataset has its own best alignment. Thus, we cannot decide on a single distance between the eyes that works for all datasets.

D. Brightness

We also did experiments with varying brightness as a means of data augmentation. The motivation behind these tests was to show more varying images to the network during training. As real scenarios may not be very well illuminated, we believe this change can improve real-world case images. Table V shows the results of this experiment across all the datasets, using the best possible alignment according to the results obtained earlier.

As we can see, the brightness variation had no beneficial effect on any of the evaluated datasets. We believe that this happens because all the datasets were made in very controlled lightning conditions, and applying variations on the training set misguides the network learning process.

E. Discussion

1) Raw Images: The results on the raw images were promising. We believe that training a model using raw images of a single dataset would achieve good results. However, this model would provide a servery lack of generalization capabilities, which does not equal something good for real case scenarios.

TABLA V
EER PER DATASET WHEN USING OR NOT BRIGHTNESS-BASED DATA AUGMENTATION.

Dataset	With Brightness Var.	Without Brightness Var.
NUAA	0.36%	0.10%
Replay Attack	4.41%	2.72%
MSU-FASD	14.03%	11.77%
OULU-NPU	13.88%	12.65%

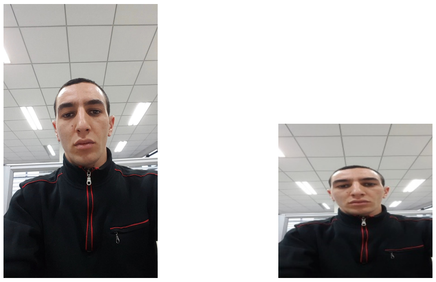

(a) Original Image
(b) Deformed Image

Fig. 3. Comparison between an original image and the same image after resize.

2) Image Stretching: Both the cropped faces and the raw images had to be resized to match the input size of the VGG16 architecture. This resize operation introduced some image deformations on the images. These deformations vary depending on the resolution of the original image and the detected bounding box size. We believe that this stretching pattern would hinder the performance in different real scenarios, despite the promising results achieved in the first two experiments. The stretching can be observed in Figure 3.

3) Brightness Variations: The brightness variations did not show any improvements in the results. The primary motivation of the data augmentation is to provide more varying scenarios in training time which would be more challenging so that the network can correctly predict a more different assortment of test images. On the other hand, the results show that introducing different brightness levels were detrimental to the model's performance. Thus, we believe that the models are still very limited in achieving good results only on the very controlled scenarios.

4) Grad-CAM Analysis: The Grad-CAM algorithm allows us to identify the regions that activated the network's responses. This way, we can identify the most critical features and regions of the image when making a classification. Figure 4 shows an example of the Grad-CAM on some images.

It is possible to see that for the attack images, the regions that activated the network's response were either around the edges of the print attacks or on the fingers of the people holding the photo. Aside from these cases, whenever an image does not activate the network, it is also classified as an attack. For authentic images, it is seen that the regions of the forehead

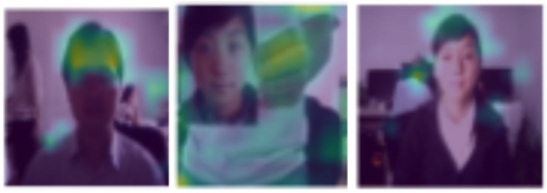


Fig. 4. Images with Grad-CAM heatmaps.

and the cheeks activate when making a correct prediction.

5) Comparison with other static methods: To better show how our results compare against other static methods, we made the table VI.

TABLA VI
EER PER DATASET WHEN USING OR NOT BRIGHTNESS IN DATA AUGMENTATION.

Dataset	Method	EER%
NUAA	LBP+Gabor Wavelets+HOG [45]	1.10%
NUAA	LBP+LPQ+HOG [46]	1.90%
NUAA	MLPQ-TOP [47]	1.10%
NUAA	**Ours**	0.10%
Replay Attack	Fine-Tuned-VGGFace [12]	8.40%
Replay Attack	DPCNN [12]	2.90%
Replay Attack	Patch Based CNN [31]	2.50%
Replay Attack	**Ours**	2.72%
MSU-MFSD	Color LBP [48]	10.80%
MSU-MFSD	GFA-CNN [49]	7.50%
MSU-MFSD	**Ours**	11.77%
OULU-NPU	DeepPixBis [33]	6.00%
OULU-NPU	FaceDs [50]	4.30%
OULU-NPU	CPqD [51]	6.90%
OULU-NPU	**Ours**	12.65%

V. CONCLUSIONS

In this work, we experimented with some preprocessing factors on some of the most well-known datasets for the face anti-spoofing problem. We aimed to achieve competitive results while also employing good practices for ensuring a good data generalization. Our approach, based on the transfer learning of a VGG16 network, achieved competitive results for similar methods that use static texture for classification. Moreover, we achieved this result while training with a fraction of the original datasets, enforcing that experiments can be run much quicker without sacrificing accuracy. We also concluded that the brightness variations as a means of data augmentation are not beneficial on the datasets. The existing datasets have very controlled lightning conditions, even though the introduction of illumination variation on the train images theoretically would be beneficial for a real case scenario. We also want to investigate the implications of training simpler networks as a means to reduce over-fitting.

ACKNOWLEDGMENT

The authors would like to thank Anna Medeiros for her invaluable insights during the making of this paper.

REFERENCES

[1] World Health Organization, "What you need to know about the new omicron covid-19 variant," 2021. [Online; accessed 09-December-2021].

[2] World Health Organization, "Listings of who's response to covid-19," 2021. [Online; accessed 09-December-2021].

[3] Centers of Disease Control and Prevention, "How to protect yourself & others," 2021. [Online; accessed 09-December-2021].

[4] M. Gomez-Barrero, P. Drozdowski, C. Rathgeb, J. Patino, M. Todisco, A. Nautsch, N. Damer, J. Priesnitz, N. Evans, and C. Busch, "Biometrics in the era of covid-19: Challenges and opportunities," *arXiv preprint arXiv:2102.09258*, 2021.

[5] Y. A. U. Rehman, L. M. Po, and M. Liu, "Deep learning for face anti-spoofing: An end-to-end approach," in *2017 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pp. 195–200, IEEE, 2017.

[6] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, *Handbook of biometric anti-spoofing: Presentation attack detection*. Springer, 2019.

[7] S. Kumar, S. Singh, and J. Kumar, "A comparative study on face spoofing attacks," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 1104–1108, IEEE, 2017.

[8] L. Sun, W. Huang, and M. Wu, "Tir/vis correlation for liveness detection in face recognition," in *International Conference on Computer Analysis of Images and Patterns*, pp. 114–121, Springer, 2011.

[9] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 436–441, IEEE, 2011.

[10] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," 6 2014.

[11] Z. Ming, M. Visani, M. M. Luqman, and J. C. Burie, "A survey on anti-spoofing methods for face recognition with rgb cameras of generic consumer devices," *arXiv*, 2020.

[12] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, "An original face anti-spoofing approach using partial convolutional neural network," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6, IEEE, 2016.

[13] B. Zinelabidine, K. Jukka, L. Li, X. Feng, and A. Hadid, "Oulunpu: a mobile face presentation attack database with real-world variations," in *Proc. IEEE Int. Conf. on Identity, Security and Behavior Analysis, ISBA*, pp. 1–7, 2017.

[14] I. Chingovska, A. Anjos, and S. Marcel, "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, IEEE, 2012.

[15] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, 2015.

[16] 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*, pp. 504–517, Springer, 2010.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, p. 336–359, Oct 2019.

[18] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," in *Biometric technology for human identification*, vol. 5404, pp. 296–303, International Society for Optics and Photonics, 2004.

[19] A. Azarbayejani, T. Starner, B. Horowitz, and A. Pentland, "Visually controlled graphics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 602–605, 1993.

[20] K. Kollreider, H. Fronthaler, and J. Bigun, "Evaluating liveness by face images and the structure tensor," in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pp. 75–80, IEEE, 2005.

[21] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *2007 IEEE 11th international conference on computer vision*, pp. 1–8, IEEE, 2007.

[22] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in "liveness" assessment," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 548–558, 2007.

[23] B. Peixoto, C. Michelassi, and A. Rocha, "Face liveness detection under bad illumination conditions," in *2011 18th IEEE International Conference on Image Processing*, pp. 3557–3560, IEEE, 2011.

[24] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[25] B. S. Manjunath and W.-Y. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 18, no. 8, pp. 837–842, 1996.

[26] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, vol. 1, pp. 886–893, Ieee, 2005.

[27] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, “Lbp- top based countermeasure against face spoofing attacks,” in *Asian Conference on Computer Vision*, pp. 121–132, Springer, 2012.

[28] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, “Computationally efficient face spoofing detection with motion magnification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 105–110, 2013.

[29] A. da Silva Pinto, H. Pedrini, W. Schwartz, and A. Rocha, “Video-based face spoofing detection through visual rhythm analysis,” in *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 221–228, IEEE, 2012.

[30] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection using 3d structure recovered from a single camera,” in *2013 international conference on biometrics (ICB)*, pp. 1–6, IEEE, 2013.

[31] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, “Face anti-spoofing using patch and depth-based cnns,” in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 319–328, IEEE, 2017.

[32] K. Patel, H. Han, and A. K. Jain, “Cross-database face antispoofing with robust feature representation,” in *Chinese Conference on Biometric Recognition*, pp. 611–619, Springer, 2016.

[33] A. George and S. Marcel, “Deep pixel-wise binary supervision for face presentation attack detection,” in *2019 International Conference on Biometrics (ICB)*, pp. 1–8, IEEE, 2019.

[34] Z. Wang, Z. Wang, Z. Yu, W. Deng, J. Li, T. Gao, and Z. Wang, “Domain generalization via shuffled style assembly for face anti-spoofing,” 2022.

[35] L. Zhou, J. Luo, X. Gao, W. Li, B. Lei, and J. Leng, “Selective domain-invariant feature alignment network for face anti-spoofing,” *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 5352–5365, 2021.

[36] D. E. King, “Dlib-ml: A machine learning toolkit,” *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[37] X. Tan, F. Song, Z. H. Zhou, and S. Chen, “Enhanced pictorial structures for precise eye localization under uncontrolled conditions,” *2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 1621–1628, 2009.

[38] A. N. Avanaki, “Exact global histogram specification optimized for structural similarity,” *Optical Review*, vol. 16, no. 6, pp. 613–621, 2009.

[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. 2016.

[40] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “VGGFace2: A dataset for recognising faces across pose and age,” *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, pp. 67–74, 2018.

[41] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[42] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[43]

[44] A. Ilugbusi and A. Adetunmbi, “Development of a multi-intance fingerprint based authentication system,” pp. 1–9, 10 2017.

[45] J. Määttä, A. Hadid, and M. Pietikäinen, “Face spoofing detection from single images using texture and local shape analysis,” *IET biometrics*, vol. 1, no. 1, pp. 3–10, 2012.

[46] J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection with component dependent descriptor,” in *2013 International Conference on Biometrics (ICB)*, pp. 1–6, IEEE, 2013.

[47] S. R. Arashloo, J. Kittler, and W. Christmas, “Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2396–2407, 2015.

[48] Z. Boulkenafet, J. Komulainen, and A. Hadid, “Face anti-spoofing based on color texture analysis,” in *2015 IEEE international conference on image processing (ICIP)*, pp. 2636–2640, IEEE, 2015.

[49] X. Tu, Z. Ma, J. Zhao, G. Du, M. Xie, and J. Feng, “Learning generalizable and identity-discriminative representations for face anti-spoofing,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 5, pp. 1–19, 2020.

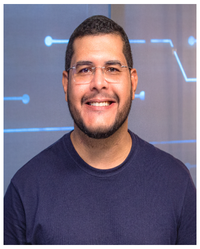
[50] A. Jourabloo, Y. Liu, and X. Liu, “Face de-spoofing: Anti-spoofing via noise modeling,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11217 LNCS, pp. 297–315, 2018.

[51] Z. Boulkenafet, J. Komulainen, Z. Akhtar, A. Benlamoudi, D. Samai, S. E. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, L. Qin, *et al.*, “A competition on generalized software-based face presentation attack detection in mobile scenarios,” in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 688–696, IEEE, 2017.

Sandoval Verissimo is graduated in Computer Engineering by the Federal University of Paraiba (UFPB), Brazil. He is currently a Master candidate by the same university, focusing on Computer Vision and Deep Learning.

Guilherme Gadelha is graduated in Computer Science by the Federal University of Campina Grande (UFCG), Brazil. He has a Master’s Degree in Computer Science and is currently a Ph.D. candidate by the same university, focusing on Computer Vision and Deep Learning.

Leonardo Vidal Batista received the M.Sc degree in Electrical Engineering from the Pontifical Catholic University of Rio de Janeiro, Brazil, in 1993, and the Ph.D. degree from the Federal University of Campina Grande, Brazil, in 2002. From 1990 to 1993 he worked at the Scientific Center of IBM Brazil. He is now a Professor at the Computer Systems Department of the Federal University of Paraíba, Brazil. His research fields include computer vision and artificial intelligence.



Fabio Falcão has a Master Degree in Informatics Engineering focused in Data Science and Computer Forensics from Universidade de Coimbra, Portugal. Undergraduate and graduate professor in Machine Learning & IoT and Big Data MBA. CEO of IARIS Tech, a company specializing in AI solutions.

João Janduy is M.Sc. degree in Computer Science from the Federal University of Paraíba and Ph.D. in Computer Science from the Federal University of Campina Grande, specializing in Biometrics, focusing on Fingerprints and Facial Recognition.